KU LEUVEN

FACULTY OF ECONOMICS
AND BUSINESS

Nonparametric analysis of production and consumption behavior

Dissertation presented to obtain the degree of Doctor in Economics

by

Veerle Hennebel

Number 523 Year 2016

Daar de proefschriften in de reeks van de Faculteit Economische en Toegepaste Economische Wetenschappen het persoonlijk werk zijn van hun auteurs, zijn alleen deze laatsten daarvoor verantwoordelijk.

Doctoral committee

Supervisor:	Prof. dr. Laurens Cherchye
Co-supervisor:	Prof. dr. Bram De Rock
Members:	Prof. dr. Kristof De Witte
	Prof. dr. Sergio Perelman
	Prof. dr. Tom Van Puyenbroeck
Chairman:	Prof. dr. Geert Dhaene

Acknowledgements

May 2016 has been a truly special month. My life has changed in many respects the last couple of years and this month was the culmination of those years. Now that I am looking back at the last couple of years, it is a pleasure to thank the people that helped me along my way.

First of all, I want to express my gratitude to my supervisor Prof. Laurens Cherchye, for his continuous support and his faith in me. He has been a great supervisor and his guidance and encouragement have been an enormous help. Furthermore, this PhD would not have been possible without my co-supervisor Prof. Bram De Rock. I want to thank him for his many helpful suggestions and the support throughout the last couple of year. Moreover, I want to thank my co-authors Sam and Marijn, I sincerely enjoyed working on our joint projects.

I am truly honored that Prof. Kristof De Witte, Prof. Sergio Perelman and Prof. Tom Van Puyenbroeck were willing to accept the invitation to become a member of my PhD jury. I am grateful for the time they have invested in this dissertation. As experts in their fields, their comments greatly improved my PhD.

I cherish the years that I have been working on my PhD and this is also the credit of my colleagues at the Subfaculteit Economie en Bedrijfswetenschappen Kulak. Your company during soup, lunch and coffee breaks means a lot to me and has made the years I spent at Kulak very enjoyable. Special thanks go to Joeri, Sam, Pieter Jan, Khushboo and Jenke for sharing an office with me.

Last, but not least, I want to thank my family and friends. Thank you to my parents, Miriam and Stefaan, you have made me the person I am today. I am lucky to have such great parents and I am very grateful for your care and continuous support. Thank you also to my parents-in-law, Ignas and Marleen, for welcoming me in such a warm family. Thank you Kristof, Yentel, Andres, Evelien, Jolien, Jennis, Lise, Nathalie and Bram, for the great times that we have together. Most of all I want to thank my husband Mathias, for the many beautiful moments we share. You have been a great support and are taking very good care of me. With you by my side, I have faith that the future will be bright.

Veerle

Contents

List of Tables

List of Figures

Part I

General introduction

In my dissertation I extend and apply tools that are designed to test consistency of observed behavior with theoretically optimizing behavior. The first three chapters of my dissertation will treat production behavior, the fourth chapter treats consumption behavior. Although production and consumption behavior are two different fields of study, there is a close link. The motivation to combine the study of production and consumption behavior in this dissertation is twofold. First, there is a conceptual analogy between production and consumption behavior. On the one hand, a producer uses inputs to produce outputs. On the other hand, a consumer could be seen as if he uses goods to 'produce' utility. The level of utility, produced by a bundle of goods, will then depend on the preferences of the consumer in question. Just as an optimizing producer is called efficient, an optimizing consumer is called rational. The main difference between the two fields pertains to the observability of the outputs of the production process. In the production setting the output levels are usually observed, which leads to a powerful analysis of efficient production behavior. In the consumption setting, an additional challenge is that it is practically impossible to observe utility levels. However, as we will explain later, the observed choices of consumers implicitly reveal information on the preferences of the consumers, which will allow us to analyze whether the observed behavior is rational or not. This approach is known in the literature as the revealed preference approach to consumer demand.

A second motivation to combine the study of production and consumption in this dissertation is that production and consumption are interrelated in a household context. In the theory of household production (Becker (1965)), households are both producers and consumers of goods. Households combine time and market goods to produce domestic commodities. For example, home cooked meals are produced by means of food and cooking time. The domestically produced commodities then directly contribute to the utility of the household members. In the fourth chapter we will integrate production in a consumption context. To our knowledge, Chapter 4 is the first study that explicitly integrates household production in a revealed preference analysis of household consumption.

Both in the context of production and consumption, there is a major interest in testing consistency of observed behavior with theoretically optimizing behavior and in quantifying deviations from optimizing behavior. In the context of consumption the basic assumption of the standard (neoclassical) model of consumer behavior is that people make rational choices. In particular, a rational consumer chooses a bundle of goods that maximizes his utility, subject to a budget constraint. Since rationality of the economic agent forms the starting point of choice theory, it is crucial to investigate whether the assumption of rationality makes sense in a particular setting. At this point it might be worth to note that, given consistency with the behavioral model, other issues concerning the behavioral model can be addressed. In particular, Varian (1982) mentions the recovery of consumer preferences and the prediction of behavior in other economic situations (such as other price regimes). However, the focus of this dissertation will be the testing of consistency.

In the context of production, the main interest is to quantify how close the observed production behavior is to optimizing behavior. This is called the degree of 'efficiency' of the observed behavior. Efficiency analysis of production activities is of major interest both in the public and private sector. If inefficiencies are detected, the comparison with efficient peers can then provide valuable information to remedy the observed inefficiencies. By producing more efficient, managers can get more results with the same means.

I will follow a nonparametric approach to production and consumption analysis. The attractive feature of a nonparametric approach - in contrast to parametric methods - is that there is no need to impose (unverifiable) functional representations to analyze observed behavior. For example, parametric methods to analyze efficient production behavior rely on a functional specification of the production process. However, the effects of inefficiency could be confounded with those of a misspecification of the functional form (Fried et al. (2008)). This issue is avoided by using a nonparametric method. Similarly, parametric methods in the context of consumption behavior require a functional specification of consumer demand. Consequently, in a parametric framework, one always tests a joint hypothesis: whatever one wants to test, plus the additional assumptions concerning the functional form (Varian (1983)). An attractive feature of a nonparametric approach is that it rules out the issues corresponding to this joint hypothesis: a nonparametric approach - such as revealed preference analysis - can provide a test of rationality, without additional assumptions concerning a functional form.

To summarize, the first three chapters of my dissertation nonparametrically analyze production behavior, the fourth chapter integrates household production in a nonparametric analysis of consumption. Please note that the chapters of this dissertation were originally written as separate papers. We have tried to limit the repetition of arguments in this dissertation, but there might be some overlap, to improve exposition of the ideas and the fluent reading of the chapters. Moreover, it is possible that different chapters use different notation. In line with the structure of my dissertation, the remainder of the introduction first provides some background on nonparametric production analysis and discusses our contributions in production analysis. Finally, we come back to consumption analysis, to discuss revealed preference analysis and the integration of household production in a consumption framework.

Production behavior

Data envelopment analysis (DEA) A well-known approach to evaluate the efficiency of production activities is Data Envelopment Analysis (DEA). DEA is nonparametric in nature: there is no need to assume a specific functional representation of the production technology. Instead, DEA is a 'data oriented' method, which estimates a 'best practice' frontier, within a set of comparable production units. Furthermore, the method is able to measure the level of efficiency of the units, which are not situated on the best practice frontier. In particular, DEA reconstructs the production possibility set on the basis of the observed input-output combinations, using standard production axioms. A unit's efficiency is then measured as the distance from its input-output combination to the frontier of this production possibility set.

Besides DEA, a well-known econometric technique to analyze efficiency is stochastic frontier analysis (SFA). A strength of SFA is that the method takes into account noise, such as measurement error, which is largely ignored by standard DEA methods. In return, SFA methods assume a given functional form for the relationship between inputs and an outputs and need to choose a distribution for the noise and for inefficiency. A good reference is the handbook Coelli et al. (2005), which extensively discusses both DEA and SFA and the strengths and limitations of both methods. Moreover, Cherchye et al. (2016) discuss how both streams of literature deal with empirical challenges.

In this dissertation, we focus on the nonparametric method DEA. Building on the work of Farrell (1957), the seminal paper 'Measuring the efficiency of decision making units' by Charnes et al. (1978) introduced DEA in its present form. The term Decision Making Unit (DMU) refers to a productive organization, which converts multiple inputs into multiple outputs. The definition of a DMU is flexible and DEA is widely applicable for efficiency estimation both in the public and private sector.

We refer to Färe et al. (1994), Cooper et al. (2007) and Cook and Seiford (2009) for overviews of the major developments in DEA over the past three decades. As Cook and Seiford (2009) indicate, there has been an impressive growth both in theoretical developments and applications of the ideas to practical situations. In my dissertation, I hope to contribute in both respects.

Economic perspective on DEA Afriat (1972), Hanoch and Rothschild (1972), Diewert and Parkan (1983) and Varian (1984), among others, have advocated a 'behavioral' nonparametric approach to analyzing producer behavior. This approach starts from a behavioral model of optimizing behavior. Depending on the context, the behavioral approach often assumes that producers pursue either profit maximization or cost minimization.

In Chapter 1, we discuss the 'economic' perspective on DEA: we start from a clear specification of the production-behavioral models and impose the least structure on the production possibilities. The economic perspective is opposed to the conventional 'axiomatic' DEA approach for reconstructing production possibilities. However, some production properties are difficult to verify in practice and one should be careful to impose production properties that cannot be justified in a convincing way. We argue in favor of carefully checking the empirical validity of axioms and for investigating the sensitivity of the results with respect to these axioms if it is difficult to verify them empirically.

An axiom that is often questioned, is convexity of the production possibility set. Free disposal hull models (FDH), initiated by Deprins et al. (1984) and Tulkens (1993), assume that only the observed DMUs make up the frontier, and not convex combinations of those units (such as in standard DEA models). We discuss in Chapter 1 how duality relationships between the behavioral and the axiomatic approach may justify the use of convexity, at least in particular settings. In this dissertation, we therefore work in a setting with relaxed convexity assumptions.

Starting from the behavioral perspective of DEA, several extensions can be made to strengthen the efficiency analysis and to enhance realism. This will be the subject of Chapter 2 and 3 in my dissertation. These chapters consist both of a methodological and an empirical part.

Additional structure As a first extension, it may be useful to add structure to the transformation process from inputs to outputs. Traditional DEA methods typically treat DMUs' production processes as a black box and use no information at all on how the inputs and outputs are exactly linked to each other. However, there is in reality often information available on the link between inputs and outputs. A growing stream of DEA literature has contributed to opening the black box of efficiency measurement. For example, the network DEA model has been built around the concept of subtechnologies (see Färe and Grosskopf (2000), Färe et al. (2007) and Cook and Zhu (2014)). Another example is Cherchye et al. (2013), who introduce a multi-output methodology, using the notion of output-specific production technologies. Cherchye et al. (2013) show that including information on the allocation of inputs to outputs considerably strengthens the analysis: the introduced methodology has more power to identify inefficient production behavior. A crucial difference between the methods above is that Cherchye et al. (2013) deal with joint inputs in the production model, which are not considered in the existing network DEA models. In particular, Cherchye et al. (2013) distinguish between output-specific inputs, which can be allocated to particular outputs, and joint inputs, which simultaneously benefit the production of multiple outputs. By including joint inputs in the analysis, the authors model the presence of economies of scope. Economies of scope occur if the average production cost decreases when the number of outputs increases (Baumol et al. (1982)), which forms a prime economic motivation for simultaneously producing multiple outputs.

In Chapter 2, we build on the multi-output methodology of Cherchye et al. (2013) and propose an extension that quantifies possible efficiency gains by reallocating inputs over outputs. In particular, we introduce a measure of coordination efficiency, which captures these efficiency gains. Furthermore, we illustrate the proposed methodology by studying the education and research conducted at US universities. We believe that university performance forms an interesting application area, because universities typically have a two-fold assignment, i.e. education and research. Interestingly, we can distinguish between expenses that can be allocated directly to the research and eduction divisions and expenses that have a joint nature. To be concrete, the output-specific inputs in our application are university expenses that are clearly directed towards either education or research. Next, joint inputs contain expenditures related to "public" services like libraries, museums, media, technology and administration. Given our specific methodological contribution, a primal focus in Chapter 2 will be on the (efficient) allocation of university budgets over education and research outputs.

Additional technological information Second, we may impose additional technological information, for example on returns to scale. Returns to scale have been widely studied within the framework of DEA (see e.g. Banker et al. (2004) for an overview). In Chapter 3 we extend Cherchye et al. (2013) by including alternative returns to scale assumptions in the methodology. If there is no information available on the nature of the returns to scale, one usually assumes variable returns to scale. By contrast, if there is a plausible argumentation that decreasing, increasing or constant returns to scale prevail, one could impose this returns to scale assumption to obtain a methodology that has more power to identify inefficient behavior. To formally include alternative returns to scale assumptions in the methodology, we build on the work of Petersen (1990) and Bogetoft (1996). However, our final intention in Chapter 3 is not to impose a particular returns to scale assumption, but to estimate the most appropriate returns to scale assumption for each output. To estimate returns to scale, we follow a method discussed by Podinovski (2004a) and Podinovski (2004b). An interesting feature of the methodology in Chapter 3 is that it allows for imposing and estimating returns-to-scale that are specific to individual outputs. This leads to a more realistic model, in which different outputs can experience different impacts relative to proportional changes in the related inputs. This in contrast to traditional DEA models, which assume that the returns to scale classification applies to the entire input-output bundle. When estimating returns to scale, it is crucial to control for environmental factors, that may influence the ability to convert inputs into outputs. We therefore advocate a methodology that explicitly takes into account environmental heterogeneity. Since the relevant environmental factors might differ from output to output, we allow the environment to be output-specific. Finally, since the estimation of returns to scale is sensitive to outliers, we combine our methodology with the robust order-m method (Daraio and Simar (2007a)). Intuitively, the robust method mitigates the impact of outliers by resampling from the original sample.

In Chapter 3, we study output-specific returns to scale in prisons in England and Wales. We use publicly available data, provided by the Ministry of Justice. In the context of prisons, we argue that it is crucial to consider output-specific returns to scale. We take three output objectives into account. Naturally, we consider the incarceration of convicts as one of the main outputs of a prison. Besides incarcerating convicts, we consider in our study also qualitative outputs including the provision of a humane prison environment and successful reintegration. Depending on the output in question, the opinions on the optimal scale size strongly differ. The usual motivation for large prisons is a reduction of the cost-per-place. Meanwhile, opponents fear a low quality of life and little prospects for reintegration in large-scale prisons. Chapter 3 therefore focuses on what we call the potential 'prison size dilemma'. Our multi-output methodology allows us to empirically test whether the optimal scale size of prisons differs when the focus is either on costs-per-place, quality of life in prison or successful reintegration.

To our knowledge, we posit an original estimation strategy that adequately models the multidimensional prison production process. The advocated methodology is tailored to all specificities of the prison production process and enables us to meaningfully answer the prison size dilemma, by using publicly available data. Moreover, we discuss in detail how public policy makers can further refine the analysis by adding information on the allocation of expenses to particular outputs.

Consumption behavior

In the final chapter of my doctoral dissertation, we contribute to the revealed preference literature to study household behavior. In particular, we integrate household production in a revealed preference analysis of household consumption. Before discussing household production, we first provide some background on revealed preference methodology, which is a method to analyze consumption choices.

Revealed preference Revealed preference methodology was initiated by Samuelson (1938, 1948), Houthakker (1950), Afriat (1967), Diewert (1973) and Varian (1982). The method starts from a data set of observations on consumer behavior - with information on quantities and prices - and constructs 'revealed preference relations', indicating which bundles of goods are preferred over others. Varian (1982) argues that the basic behavioral hypothesis in the context of consumption is that the consumer chooses a bundle of goods that is preferred to all other bundles that he can afford. Stated more formally, a rational consumer chooses the bundle of goods that maximizes its utility, subject to a budget constraint. Varian (1982) then shows that a bundle is utility maximizing subject to its budget constraint if and only if it is expenditure minimizing over all the revealed preferred bundles. The latter is easily verified and is known as the Generalized Axiom of Revealed Preference (GARP). A major advantage of revealed preference methodology is that it allows us to test rationality, without making (unverifiable) assumptions about a functional specification of the preferences. Revealed preference methodology thus provides a nonparametric alternative to the standard parametric approach to consumer demand.

Households with children In Chapter 4 we study the consumption and time use of a sample of Dutch couples with children. There is some discussion in the literature on the role of children in the decision making process of the household (e.g. Dauphin et al. (2011)). This discussion is particularly relevant for older children (age 16 and older), however it is less debatable that younger children are bystanders in the household. To model the presence of (younger) children in the household, children are often treated as a public good in the household (e.g. Blundell et al. (2005) and Cherchye et al. (2012)). Parents care for the well-being of the children and can generate (or 'produce') well-being for the children by spending time together and by spending money on the commodities of the children. This is the approach we will follow. In particular, we look into the time that parents invest into their children, such as bathing, dressing, playing, reading stories, going with the child to the doctor and taking the child to school or hobbies.

Household production and process benefits In a household labor supply model, households need to decide on consumption and time use. With respect to time use, there is typically a trade-off between hours of leisure and income as a result of hours of market labor. Besides leisure and market work, a third option is that household members also allocate time to some domestic production activity. Just some examples of domestic production are cleaning, cooking or caring for children. Becker (1965) and Gronau (1977) assume that households can produce domestic commodities by allocating time and resources to the corresponding activity. As mentioned before, we will focus in Chapter 4 on one specific domestic commodity: child well-being. Child well-being is treated as a domestic commodity, which is produced by the parents of the household, by a combination of time spent with children and consumption goods allocated to children.

It seems reasonable that time spent on a particular production activity generates both indirect utility, through the commodity which is produced, and direct utility in terms of leisure. For example, child care time may also be perceived as leisure for the caring parent. This is captured by the notion of 'process benefits': parents may enjoy the process of caring for the children. Process benefits are a form of joint production. Pollak and Wachter (1975) argue that when time is an input in the household production process, joint production is the rule and not an exception. Although allowing for joint production complicates the estimation of the household production function - see Pollak and Wachter (1975) for a discussion - it is perfectly possible to analyze the allocation of goods and time among household activities. Graham and Green (1984), Kerkhofs and Kooreman (2003) and Gørtz (2011) extend the model of Gronau (1977) to take process benefits into account. However, these papers focus on estimating a parametric version of the model, with and without process benefits.

By contrast, in Chapter 4, we analyze process benefits using revealed preference methodology. Doing so, there is no need to specify a functional form for the utility function or the household production function. Moreover, the output of the home production process need not be observed. This makes our methodology particularly useful to investigate non-marketable goods such as child well-being.

Part II

Production behavior

Chapter 1

The economic meaning of Data Envelopment Analysis: a ‘behavioral’ perspective

Abstract

We reconsider the motivation of Data Envelopment Analysis (DEA), the non-parametric technique that is widely employed for analyzing productive efficiency in academia, the private sector and the public sector. We first argue that the conventional engineering motivation of DEA can be problematic since it often builds on unverifiable production axioms. We then provide a dual viewpoint and highlight the 'behavioral' interpretation of DEA models. We start from a specification of the production objectives while imposing minimal structure on the production possibilities, and construct tools to meaningfully quantify deviations of observed producer behavior from optimizing behavior. This brings to light the economic meaning of DEA, provides guidelines for selecting the appropriate model in practical research settings, and prepares the ground for instituting new DEA models. We hope that our insights will contribute to the further dissemination of DEA, and stimulate public sector applications of DEA that build on its behavioral interpretation.[1]

[1] This chapter is based on joint work with Laurens Cherchye (KU Leuven) and Bram De Rock (ULB). I refer to the paper of Cherchye et al. (2014b), which has been published in Socio-economic planning sciences.

1.1 Introduction

The public sector is increasingly interested in the productive efficiency of its entities. For instance, Coelli et al. (2003) extensively discuss the relevance of efficiency evaluations for regulated sectors. More generally, the growing number of empirical applications suggests that productive efficiency analysis is of key interest for many sectors such as academia, the business community and government institutions; see, e.g., Gattoufi et al. (2004) and Emrouznejad et al. (2008) for overviews. This observation calls for well-established empirical tools that are specially tailored for testing consistency of observed behavior with (theoretical) optimizing behavior, and for quantifying deviations from optimization (or 'inefficiencies').

Afriat (1972), Hanoch and Rothschild (1972), Diewert and Parkan (1983) and Varian (1984), among others, have advocated a 'behavioral' non-parametric approach to analyzing producer behavior. This approach starts from a behavioral model of optimizing/efficient behavior and allows for testing implications of micro-economic theory directly on the data. That is, one does not need a functional representation of the production technology, and so one can minimize the risk of erroneously rejecting optimizing producer behavior due to an erroneous parametric specification of the (typically unknown) technology. This is particularly convenient, since economic theory does in general not imply a particular functional form and reliable specification tests are not available in many cases.

Non-parametric efficiency analysis is increasingly applied for measuring the degree of 'efficiency' of observed producer behavior, most commonly under the label 'Data Envelopment Analysis' (DEA; after Charnes et al. (1978)).[2] DEA models are conventionally motivated from 'engineering' information, e.g. pertaining to the prevalent returns-to-scale or the marginal rates of input substitution/output transformation. Still, such engineering information is mostly difficult to verify in practice. In fact, imposing production properties that cannot be justified in a convincing way seems to conflict directly with the very nature of non-parametric analysis, which is often credited for imposing minimal structure on the research setting under investigation. This consideration is particularly relevant for DEA

[2] See Färe and Grosskopf (1994), Cooper et al. (2007), Fried et al. (2008), and Cook and Seiford (2009) for extensive surveys of DEA models.

evaluations of the public sector, which are usually characterized by minimal information on the nature of production possibilities.

In this chapter, we adopt an 'economic' (as opposed to 'engineering') perspective on DEA: we start from a clear specification of the production-behavioral models and use minimal (non-verifiable) engineering information. Our insights re-interpret DEA efficiency measures as measures for violations of economically optimizing behavior. To keep our exposition simple, we mainly focus on profit maximizing and cost minimizing behavior. However, as we will indicate, our insights readily extend towards alternative production-behavioral models. By making explicit this economic motivation of DEA, we hope to contribute to its further dissemination and to stimulate public sector applications of DEA that build on its behavioral interpretation.

We note at the outset that our discussion bears some analogy to that in Varian (1990) and Färe and Grosskopf (1995), where a similar interpretation of DEA efficiency measures is (implicitly) advocated. Unfortunately, although these ideas have some clear advantages, they are only minimally used in the applied DEA literature; see, e.g., Cherchye et al. (2008, 2013, 2014c) for some applications that demonstrate the advantages of the behavioral perspective of DEA. If only for that reason, it seems useful to set out methodological guidelines for economically meaningful applications of DEA. Furthermore, our discussion includes a number of insights that have not yet been articulated in the literature, and prepares the ground for instituting new DEA models depending on the production-behavioral model that is subject to testing.

The remainder of this chapter unfolds as follows. In Section 2 we briefly review the conventional 'axiomatic' DEA approach for reconstructing production possibilities. Section 3 is concerned with non-parametric economic efficiency analysis, following the perspective of Afriat (1972), Hanoch and Rothschild (1972), Diewert and Parkan (1983) and Varian (1984). Section 4 bridges the gap between the seemingly distinct viewpoints adopted in Sections 2 and 3, and brings to light the economic meaning of DEA. Section 5 presents an empirical application on efficiency in academia and illustrates the relevance of an appropriate DEA model. Section 6, finally, reproduces the main insights and provides some concluding discussion.

1.2 Reconstructing production possibilities: an axiomatic approach

A producer creates outputs from various combinations of inputs (factors of production). To study producer choices we need a convenient way to summarize the production possibilities, i.e. which inputs and outputs are *technologically feasible*. The set of all technologically feasible input-output combinations is called the *production possibility set*.

To formally represent that set, we denote by $z = (z^1, \ldots, z^q) \in \mathbb{R}^q$ a (non-zero) netput vector with z^j the value of netput commodity *j*. Positive components of *z* represent outputs and negative components represent inputs. Throughout we assume that the vector *z* captures at least one input and at least one output, and that all producers use the same commodities as inputs and produce the same outputs. The production technology is represented by the (non-empty and closed) production possibility set

$$T = \{z \in \mathbb{R}^q | \text{ netput } z \text{ is technically feasible}\}. \tag{1.1}$$

If we make the explicit distinction between input and output vectors, we use $z = (-x, y)$ with $x \in \mathbb{R}^l_+$ the input vector and $y \in \mathbb{R}^m_+$ the output vector ($q = l + m$). Then, the set T can be decomposed into input requirement sets

$$L_T(y) = \left\{x \in \mathbb{R}^l_+ |(-x, y) \in T\right\}, \tag{1.2}$$

which contain all input vectors x that can produce the output vector y.

Production axioms. The true production possibility set T (or the input requirement set $L_T(\cdot)$) is usually not observed. Therefore the DEA-type axiomatic approach typically approximates the unobserved set T by an empirical production set that is constructed from a set of observed producers. We represent each observed producer s by the netput vector $z_s = (-x_s, y_s)$, with $s \in S = \{1, \ldots, |S|\}$, for S the set of observed producers. To construct the empirical approximation of T, we will consider the production axioms 1.1-1.4.[3]

[3] In a theoretical framework, Shepard (1970) provides a comprehensive list of production axioms (including ours), which we do not intend to fully review. Other axioms presented in the DEA literature (see, e.g., Färe et al.

Axiom 1.1 (Inclusion of observations)**.** $\forall s \in S : (-x_s, y_s) \in T.$

This axiom says that all observed netput vectors are technologically feasible and thus that they should belong to the (unobserved) production set T. This is really an empirical postulate rather than a production postulate. It makes that we exclude empirical phenomena such as measurement error or outlier behavior.[4]

Axiom 1.2 (Monotonicity)**.** If $z \in T$ and $z' \leq z$ then $z' \in T.$

Monotonicity, sometimes also referred to as 'strong (or free) disposability' of inputs and outputs, implies that the producer can always costlessly dispose unwanted inputs and/or outputs. That is, more inputs cannot lead to producing less outputs and producing less outputs cannot lead to using more inputs. It implies that marginal rates of substitution/transformation (between inputs, between outputs and between inputs and outputs) are nowhere negative or, in other words, there is no congestion.

Axiom 1.3 (Convexity in netput space)**.** If $z \in T$ and $z' \in T$, then $\lambda z + (1 - \lambda) z' \in T$ for all $\lambda \in [0, 1]$.

Axiom 1.4 (Convexity in input space)**.** If $x \in L_T(y)$ and $x' \in L_T(y)$, then $\lambda x + (1 - \lambda) x' \in L_T(y)$ for all $\lambda \in [0, 1]$.

Convexity in netput space entails that marginal rates of substitution/transformation (between inputs, between outputs and between inputs and outputs) are nowhere increasing. Convexity in input space, finally, is a weaker version of Axiom 1.3 and entails non-decreasing marginal rates of input substitution.

Apart from these specific production axioms, the (axiomatic) DEA approach typically builds on a 'minimal extrapolation' requirement, which says that the production set approximation should be the minimal set that is consistent with the axioms adopted; see Banker et al. (1984).

Production set approximations. Different production set approximations are obtained from different sets of axioms. First, if we impose Axioms 1.1 and 1.2, then the resulting

(1994)) are not considered because they are not instrumental to our following discussion.

[4]See, e.g., Grosskopf (1996) for extensions of DEA that weaken this assumption.

production set approximation consistent with the minimum extrapolation principle is the monotone hull of the data: $M(S)$.[5]

$$M\left(S\right)=\left\{z\in\mathbb{R}^{q}\left|z\leq z_{s}\text{ for some }s\in S\right.\right\} \tag{1.3}$$

Second, if we additionally assume convexity in the netput space (i.e. Axiom 1.3), then we get the convex monotone hull of the data: $CM(S)$.[6]

$$CM\left(S\right)=\left\{z\in\mathbb{R}^{q}|\forall s\in S:z\leq\sum_{s\in S}\lambda_{s}z_{s},\lambda_{s}\geq0\text{ and }\sum_{s\in S}\lambda_{s}=1\right\} \tag{1.4}$$

Finally, replacing Axiom 1.3 by Axiom 1.4 leads to the approximation $CIM(S)$, which corresponds to $M(S)$ with the additional property that input requirement sets are convex.[7]

$$CIM\left(S\right)=\left\{\left(-x,y\right)\in\mathbb{R}^{q}\left|\begin{array}{l}\forall s\in S:x\geq\sum_{s\in S}\lambda_{s}x_{s}\text{ and }\lambda_{s}y\leq\lambda_{s}y_{s}\\ \text{with }\lambda_{s}\geq0\text{ and }\sum_{s\in S}\lambda_{s}=1\end{array}\right.\right\} \tag{1.5}$$

Increasing stringency of the different assumptions underlying these three production set approximations implies

$$M(S)\subseteq CIM(S)\subseteq CM(S). \tag{1.6}$$

The sets $M\left(S\right)$, $CM\left(S\right)$ and $CIM\left(S\right)$ are illustrated in Figures 1.1 and 2 for respectively netput space and input space. Figure 1.1 represents these production set approximations for a situation with 3 producers that use a single input to produce a single output, i.e. $S_1=\{1,2,3\}$ and $z_s=(-x_s,y_s)\in\mathbb{R}_-\times\mathbb{R}_+$, with $s\in S$. The monotone hull of the data $M\left(S_1\right)$ is the area under the full line, while $CM(S_1)$ coincides with the area under the dotted line. Observe further that for this particular situation (with only one input and one output) $M\left(S_1\right)=CIM\left(S_1\right)$. Figure 1.2 represents the input requirement sets for a situation with 3 producers that each produce the same output with two inputs, i.e. $S_2=\{1,2,3\}$

[5] See Afriat (1972) for more discussion. Deprins et al. (1984) and Tulkens (1993) suggested this approximation in a DEA context.

[6] See Afriat (1972) for more discussion. Banker et al. (1984) proposed it in a DEA context.

[7] See Hanoch and Rothschild (1972) for more discussion. Bogetoft (1996) considers this approximation in a DEA context.

and $z_s = (-x_s, y_0) \in \mathbb{R}^2_- \times \mathbb{R}_+$, with $s \in S$. As the three producers in S_2 produce exactly the same output y_0, we get that $L_{CM(S_2)}(y_0) = L_{CIM(S_2)}(y_0)$.

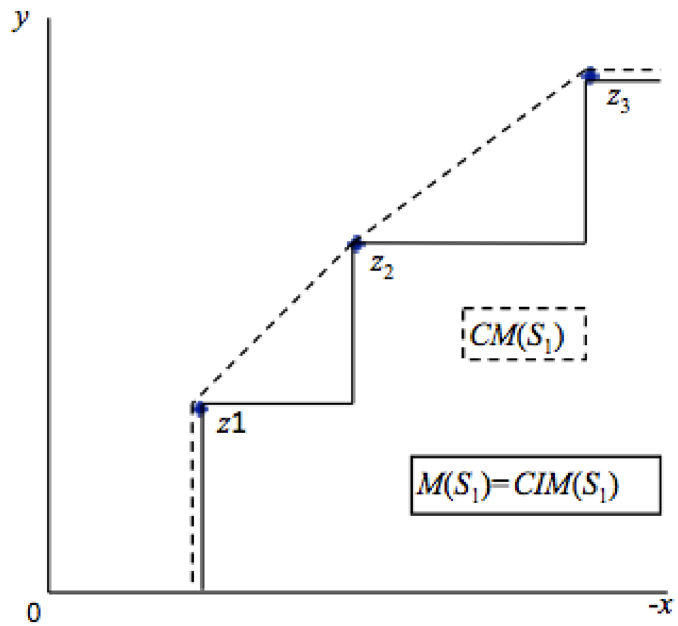


Figure 1.1: Empirical production possibility sets

From these illustrations we can conclude that production axioms directly affect the empirical production set. Hence, an important question pertains to the validity of these axioms. Unfortunately, there does not seem to exist any a priori reason why a production set should necessarily be monotone or convex. In fact, it turns out that monotonicity and convexity assumptions are problematic in many practical settings, and that reliable non-parametric specification tests are currently not available; see Cherchye and Post (2003) for an in-depth discussion. (McFadden, 1978, pp. 8-9) aptly summarizes that the common rationale for monotonicity and convexity assumptions in production theory lies in their analytical convenience rather than in their economic realism. As such a 'non-engineering' justification of DEA is recommendable, which motivates our 'behavioral' (or 'economic') perspective in the next section.

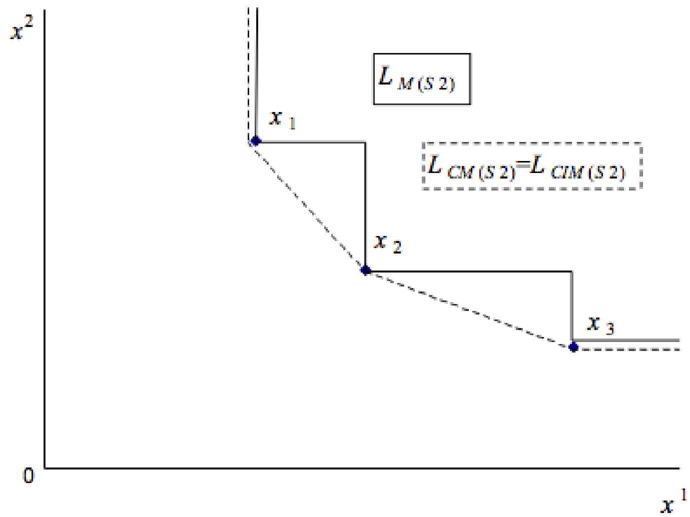


Figure 1.2: Empirical production possibility sets

1.3 Economic efficiency analysis: a non-parametric approach

While the axiomatic approach focuses on the specification of production possibilities, we now take the *dual* perspective: we start from a specification of the production objectives and impose the least structure on the production possibilities. Production objectives vary in different situations. The most frequently maintained position is that producers pursue profit maximization. In some instances, however, cost minimization for given output might seem a more reasonable assumption. For instance, when the producer is a price taker in input markets but operates in regulated output markets (as is often the case for public agencies).

In the following, we focus on profit and cost efficiency analysis of producer k ($\in S$), i.e. the producer associated with netput choice $z_k = (-x_k, y_k)$. In the first subsection, we assume that profit efficiency and cost efficiency is evaluated at (non-zero) price vectors $p_k \in \mathbb{R}^q_+$ and $w_k \in \mathbb{R}^l_+$, respectively. Note that the first part of the netput price vector p_k corresponds to the input price vector w_k, such that $p_k = (w_k, \ldots)$. In the second subsection we deal with the setting in which this price information is not available.

1.3.1 Cost and profit efficiency with price information

The minimum cost that could have been achieved by producer *k* (i.e. the producer associated with input choice x_k) when producing y_k is[8]

$$c_T(z_k, w_k) = \min_{x \in L_T(y_k)} x w_k. \tag{1.7}$$

We say that producer *k* acts cost efficient if the observed cost equals the minimum cost (i.e $x_k w_k = c_T(z_k, w_k)$) and cost inefficient if $c_T(z_k, w_k)$ is below $x_k w_k$.

Similarly, the maximum attainable profit at p_k is defined as

$$\pi_T(z_k, p_k) = \max_{z \in T} z p_k. \tag{1.8}$$

Again, profit efficiency (resp. inefficiency) is achieved by producer *k* when $z_k p_k = \pi_T(z_k, p_k)$ (resp. $z_k p_k < \pi_T(z_k, p_k)$).

Inefficient production behavior is often observed in practice and can can have different interpretations; see, e.g., Demsetz (1997) for an extensive discussion. Observed producer inefficiency can be interpreted in at least two ways. First, the producer optimization problem may be ill-specified. For example, the producers objective function can be erroneously defined; e.g. the objective function may not be fully linear in netputs (due to imperfect competition). Second, as the specified producer objective is typically that of producer owners, producer inefficiency can also be interpreted as would the producer owners incompletely control the producer managers (i.e. inefficiency due to agency problems). Both explanations instantiate the need for economic efficiency *measures*, to serve either as indicators of 'economic significance' of specification errors (see Varian (1990) or as 'performance' indicators (and possible monitoring instruments for producer owners; see Bogetoft (2004)).

Intuitively, meaningful efficiency measures give us an idea about how 'close' observed behavior is to optimizing behavior. In general, a reasonable measure of 'closeness' tells us how far the producer fails to optimize the postulated objective function. For example, when the production objective is specified as profit maximization, a reasonable measure should

[8]For simplicity we assume that minimum cost (in (1.7)) and maximum profit (in (1.8)) is defined wherever needed.

capture how much additional profit the producer could have acquired if it had behaved differently.

Cost efficiency measurement. For $x_k w_k > 0$, Farrell (1957) suggests as a measure for cost efficiency the ratio of minimal to actual cost, i.e.

$$C_T(z_k, w_k) = \frac{c_T(z_k, w_k)}{x_k w_k}. \tag{1.9}$$

It is clear that $C_T(z_k, w_k) \in [0, 1]$.[9]

As discussed above, the precise specification of T is usually unknown. Therefore, the starting point within the non-parametric approach to analyzing production behavior is that a (non-empty) subset $\{(-x_s, y_s)|s \in S\} \subseteq T$ is observed (i.e. Axiom 1.1 in Section 2). In principle, one may conduct a cost efficiency analysis by replacing T by this set. This gives minimal ('necessary') non-parametric tests for economic efficiency and upper bound estimates for the degree of cost inefficiency, i.e. $C_S(z_k, w_k) \geq C_T(z_k, w_k)$ by construction (for some $k \in S$).

However, in practice additional assumptions about the set $L_T(\cdot)$ can be useful.[10] For example, Varian (1984) assumes that less output does not require more input, i.e.

Axiom 1.5 (Free output disposability)**.** If $(-x, y) \in T$ and $y' \leq y$, then $(-x, y') \in T$.

Axiom 1.5 is a weaker version of the monotonicity Axiom 1.2. We note that our below reasoning is easily extended to accommodate for alternative assumptions regarding output disposability (like those considered in Färe et al. (1994)).

As in the previous section, we can then again obtain an approximation of the production possibility set. Axiom 1.1 and Axiom 1.5, combined with the minimal extrapolation

[9]This measure is not defined for $x_k w_k = 0$. Given that $x \in \mathbb{R}^l_+$ for all $(-x, y_k) \in T$ and $w_k \in \mathbb{R}^l_+$ we have $x_k w_k = c_T(z_k, w_k) = 0$ in that case. That is, cost efficiency is attained, and we can assign a cost efficiency value of unity to producer *k*. To keep the exposition simple we abstract from this case in the following.

[10]This enlarges the set of possible comparison partners. Otherwise, cost efficiency analysis, for example, could only compare the cost level of the evaluated producer to that of other observed producers that produce exactly the same output vector, of which the number is usually very small. However, it is worth emphasizing that cost efficiency analysis is possible even when only using Axiom 1.1. An insightful discussion of this point is given by Tulkens and Vanden Eeckaut (1999).

requirement, leads to

$$OM\left(S\right)=\left\{\left(-x_{s},y\right)|y\leq y_{s}\text{ for some }s\in S\right\}. \tag{1.10}$$

Note that by construction $OM(S)$ is a subset of the set $M(S)$, defined in (1.3), since the latter assumes monotonicity for the set T.

When Axioms 1.1 and 1.5 are rightly conjectured, necessary tests for cost efficiency can be performed with respect to $OM\left(S\right)$ and an upper bound for the cost efficiency measure in (1.9) can be derived, i.e. $C_{OM(S)}\left(z_{k},w_{k}\right)\geq C_{T}\left(z_{k},w_{k}\right)$ for some $k\in S$.

Profit efficiency measurement. Nerlove (1965) proposed two types of measures: difference measures and ratio measures. We restrict attention to ratio profit efficiency measures, since these measures have a convenient degree interpretation.[11] In addition, ratio measures are easy to work with under limited price information (see our discussion in Section 3.2).

We need to distinguish two cases. First, for $\pi_{T}\left(z_{k},p_{k}\right)>0$ we define the degree measure

$$\Pi_{T}^{+}\left(z_{k},p_{k}\right)=\frac{z_{k}p_{k}}{\pi_{T}\left(z_{k},p_{k}\right)}. \tag{1.11}$$

Second, for $\pi_{T}\left(z_{k},p_{k}\right)\leq 0$ and $z_{k}p_{k}<0$ we define

$$\Pi_{T}^{-}\left(z_{k},p_{k}\right)=\frac{\pi_{T}\left(z_{k},p_{k}\right)}{z_{k}p_{k}}. \tag{1.12}$$

Note that in the limiting case $\pi_{T}\left(p_{k}\right)=z_{k}p_{k}=0$ profit efficiency occurs. Consequently, we can simply attribute an efficiency value of unity to producer *k* in that case, i.e. $\Pi_{T}\left(z_{k},p_{k}\right)=1$ if $\pi_{T}\left(p_{k}\right)=z_{k}p_{k}=0$. Obviously, $\Pi_{T}\left(z_{k},p_{k}\right)\in\left(-\infty,1\right]$ with a value of unity revealing profit efficiency and a value below unity capturing feasible relative profit increase.

Finally, we again have to approximate T by using the set of observed netput vectors (indexed by S). As discussed above, this could lead to several different approximations ($M(S), CIM(S)$,....). For the sake of brevity, we will make abstraction of this discussion in

[11] Within the non-parametric literature difference and ratio measures for profit efficiency have been discussed by Banker and Maindiratta (1988). Our basic insights readily extend towards difference measures; compare with Cherchye and Van Puyenbroeck (2007).

the setting of profit efficiency measurement and we will only focus on the setting that starts from the observed set of netput vectors (i.e. we only impose Axiom 1.1.)

1.3.2 Measuring shadow cost and profit inefficiency

Not only the set T but also price vectors are often imperfectly observed, or the prices that are observed may not reflect the true opportunity costs perceived by producers. In that case a *shadow price* approach can be followed, i.e. basically those prices are selected that are 'most favorable' to the observation under evaluation (see, e.g., Färe et al. (1990). Below we consider the extreme case where the evaluator only knows $p_k \in \mathbb{R}^q_+$ and $w_k \in \mathbb{R}^l_+$, while excluding the zero vector. In words, we assume that prices can take any non-negative value, but they can not all be zero simultaneously. Note that, while we exclude the case where all input and output prices are zero, we still allow for zero (shadow) prices for some input and/or output commodities.

Shadow cost efficiency. Using $OM(S) \subseteq T$ the (incomplete information) counterpart of (1.9) can be defined as (for some $k \in S$)

$$C^I_{OM(S)}\left(z_k, \mathbb{R}^l_+\right) = \max_{w \in \mathbb{R}^l_+} \left\{ \frac{c_{OM(S)}\left(z_k, w\right)}{x_k w} \left| x_k w > 0 \right. \right\}. \tag{1.13}$$

To show how one can compute this measure, we have to reformulate it. In this ratio formulation prices can be scaled without affecting the value of $C^I_{OM(S)}\left(z_k, \mathbb{R}^l_+\right)$. In fact, shadow prices as obtained within the non-parametric approach typically have a ratio interpretation only. That is, they express the value of one commodity relative to that of other commodities, but they bear no direct interpretation in terms of the absolute value of each commodity, at least not without additional price information. Thus, we can set the 'shadow' cost level of producer *k* equal to unity without losing the informational content of the corresponding (relative) shadow prices, i.e. we can use

$$C^I_{OM(S)}\left(z_k, \mathbb{R}^l_+\right) = \max_{w \in \mathbb{R}^l_+} \left\{ c_{OM(S)}\left(z_k, w\right) \left| x_k w = 1 \right. \right\}. \tag{1.14}$$

Further using definitions (1.7) and (1.10), we can equivalently reformulate (1.14) as

$$C_{OM(S)}^{I}\left(z_k, \mathbb{R}_{+}^{l}\right) = \max_{w \in \mathbb{R}_{+}^{l}, c} \left\{c \,|x_k w = 1 \text{ and } c \leq x_s w \text{ for all } s \in S \text{ for which } y_s \geq y_k\right\}. \tag{1.15}$$

This last formulation makes clear that simple linear programming tools suffice to compute $C_{OM(S)}^{I}\left(z_k, \mathbb{R}_{+}^{l}\right)$. The implicit 'benefit-of-the-doubt' pricing, i.e. the selection of most favorable (shadow) prices, is reflected in the max operator. Obviously, the index $C_{OM(S)}^{I}\left(z_k, \mathbb{R}_{+}^{l}\right) \in [0, 1]$ gives an upper bound for the ratio measure $C_T\left(z_k, w_k\right)$ under incomplete price and incomplete technology information.

Shadow profit efficiency. Similarly, we can use shadow prices to deal with incomplete price information (i.e. p_k unknown) to analyze profit efficiency. Then, the analogues of the profit efficiency measures (1.11) and (1.12) are respectively

$$\Pi_S^{+}\left(z_k, \mathbb{R}_{+}^{q}\right) = \max_{p \in \mathbb{R}_{+}^{q}} \left\{\frac{z_k p}{\pi_S\left(z_k, p\right)} \,|\pi_S\left(z_k, p\right) > 0\right\} \tag{1.16}$$

and

$$\Pi_S^{-}\left(z_k, \mathbb{R}_{+}^{q}\right) = \max_{p \in \mathbb{R}_{+}^{q}} \left\{\frac{\pi_S\left(z_k, p\right)}{z_k p} \,|\pi_S\left(z_k, p\right) \leq 0 \text{ and } z_k p < 0\right\}. \tag{1.17}$$

These measures can be re-expressed as

$$\Pi_S^{+}\left(z_k, \mathbb{R}_{+}^{q}\right) = \max_{p \in \mathbb{R}_{+}^{q}} \left\{z_k p \,|z_s p \leq 1 \text{ for all } s \in S\right\} \tag{1.18}$$

and

$$\Pi_S^{-}\left(z_k, \mathbb{R}_{+}^{q}\right) = \max_{p \in \mathbb{R}_{+}^{q}, u \in \mathbb{R}_{+}} \left\{u \,|z_k p = -1 \text{ and } z_s p \leq -u \text{ for all } s \in S\right\}. \tag{1.19}$$

The possibility of zero actual profit and non-zero maximal profit is captured in $\Pi_S^{+}\left(z_k, \mathbb{R}_{+}^{q}\right)$, while the possibility of non-zero actual profit and zero maximal profit is captured in $\Pi_S^{-}\left(z_k, \mathbb{R}_{+}^{q}\right)$. The only remaining problem occurs when producer k is profit efficient only at prices that generate a zero profit level, i.e.

$$\max_{p\in\mathbb{R}^q_+} \{z_k p \,|z_k p \geq z_s p \text{ for all } s \in S\} = \min_{p\in\mathbb{R}^q_+} \{z_k p \,|z_k p \geq z_s p \text{ for all } s \in S\} = 0. \quad (1.20)$$

Such cases can be detected using linear programming tools. Clearly, we cannot reject profit efficiency when (1.20) holds.

Consistent with the idea of benefit of the doubt weighting we propose as a profit efficiency measure

$$\Pi^I_S\left(z_k, \mathbb{R}^q_+\right) = \begin{cases} \max\left\{\Pi^+_S\left(z_k, \mathbb{R}^q_+\right), \Pi^-_S\left(z_k, \mathbb{R}^q_+\right)\right\} & \text{if (1.20) does not hold} \\ 1 & \text{if (1.20) holds} \end{cases}. \quad (1.21)$$

The index $\Pi^I_S\left(z_k, \mathbb{R}^q_+\right) \in [0,1]$ gives an upper bound for the ratio measure $\Pi_T\left(z_k, p_k\right)$ under incomplete price and incomplete technology information. Not only the mere efficiency value but also the fact whether (1.20) holds and, if (1.20) does not hold, whether $\Pi^+_S\left(z_k, \mathbb{R}^q_+\right)$ or $\Pi^-_S\left(z_k, \mathbb{R}^q_+\right)$ yields the maximum in (1.21) provides useful information, and is thus preferably considered together with the profit efficiency value. As our exposition makes clear, this information tells us whether the shadow prices that are implicitly used involve a profit, a loss or a break-even for the producer under study.

1.4 Bridging the gap: the economic meaning of DEA

The dual formulation of the linear programming problem (1.15) reveals a one-to-one relationship between the above measure for cost efficiency and the Debreu (1951)- Farrell (1957) input measure for technical efficiency.[12] Similarly, the dual problems of (1.18) and (1.19) show a relationship between the proposed measure for profit efficiency and the 'McFadden gauge' function (see McFadden (1978)). These dual interpretations bring to light the economic interpretation of DEA, which typically computes technical efficiency measures

[12] This relationship in fact illustrates the duality between cost functions and the Shephard input distance functions (Shepard (1970)), which have the same informational content as the Debreu-Farrell input technical efficiency measures. In particular, the Debreu-Farrell input measure for technical efficiency is reciprocal to the Shephard input distance function; see Debreu (1951) for more discussion.

(Debreu Farrell measures) with respect to axiomatic approximations of the production possibility set. That is, it allows for interpreting these DEA measures as measures for violations of economically optimizing behavior.

1.4.1 Cost efficiency

The dual formulation of (1.15) is (for some $k \in S$)

$$C^I_{OM(S)}\left(z_k, \mathbb{R}^l_+\right) = \min_{\kappa \in \mathbb{R}_+, \lambda_s \in \mathbb{R}_{++}} \{\kappa | \sum_{s \in S} \lambda_s x_s \leq \kappa x_k, \sum_{s \in S} \lambda_s = 1 \text{ and } \lambda_s y_s \geq \lambda_s y_k \forall s \in S\}. \tag{1.22}$$

This can equivalently be reformulated as

$$C^I_{OM(S)}\left(z_k, \mathbb{R}^l_+\right) = \min_{\kappa \in \mathbb{R}_+} \{\kappa | (-\kappa x_k, y_k) \in CIM(S)\}. \tag{1.23}$$

Hence, $C^I_{OM(S)}\left(z_k, \mathbb{R}^l_+\right)$ can be computed as the maximum equiproportionate reduction of inputs within $CIM(S)$. This is precisely the Debreu-Farrell input measure defined with respect to $CIM(S)$. The fact that this reference production set is obtained falls in line with the general result that monotonizing and convexifying input requirement sets does not interfere with the analysis of cost efficiency; see Varian (1984) for more discussion. That is, the minimum cost level remains unaffected and thus $C^I_{OM(S)}\left(z_k, \mathbb{R}^l_+\right) = C^I_{CIM(S)}\left(z_k, \mathbb{R}^l_+\right)$. Hence, minimal cost reduction is also given by the maximal equiproportionate input shrinkage factor as computed with respect to $CIM(S)$.

We illustrate our discussion by means of Figure 1.3. This continues our example introduced in Figure 1.2, but now $S_2^{'} = \{(1, \ldots, 5)\}$ and $z_s = (-x_s, y_0) \in \mathbb{R}^2_- \times \mathbb{R}_+$, with $s \in S$; i.e. we include two additional observations. The input vectors are displayed in Figure 1.3. The input requirement sets associated with different sets of axioms are the same as those in Figure 1.2.

Let us first consider economic/cost efficiency. Suppose that the relative input prices correspond to the slope of the bold iso-cost line. Under these input prices, the vector x_1 is cost minimizing. Obviously, this conclusion does not change when imposing monotonicity and/or convexity on the input possibilities. The same result applies for measures of cost efficiency. For example, for the vectors x_4 and x_5 the associated cost efficiency ratios equal $0x_{4'}/0x_4$ and $0x_{5'}/0x_5$, respectively; monotonicity and convexity assumptions do not alter these results.

Next, turn to the situation of incomplete price information. From (1.23), an upper bound approximation for the cost efficiency measure is then provided by the Debreu-Farrell input measure as computed with respect to the convexified and monotonized input requirement set. The resulting value equals $0x_{4''}$ / $0x_4$ for x_4 and $0x_{5''}/0x_5$ for x_5. The upper bound interpretation is immediate: $0x_{4''}/0x_4 > 0x_{4'}/0x_4$ and $0x_{5''}/0x_5 > 0x_{5'}/0x_5$. Further, cost efficiency is achieved by x_2 and x_3; both vectors meet the necessary condition for cost minimization under the (minimal) information that is available about technology and prices.

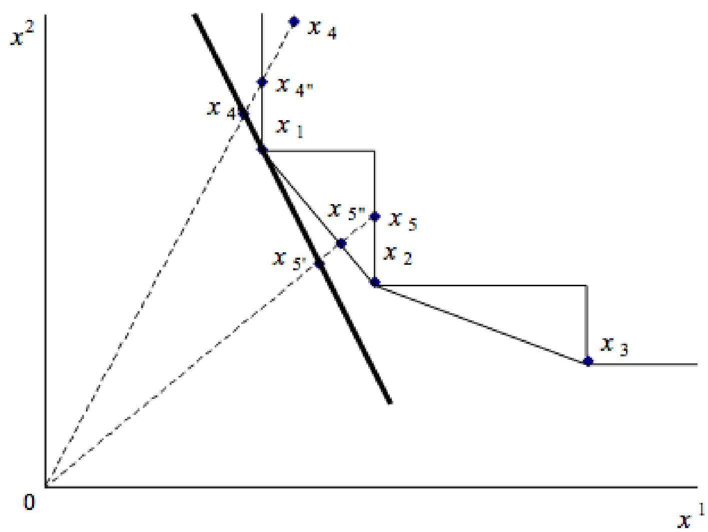


Figure 1.3: The economic meaning of DEA: cost efficiency

This example illustrates that DEA measures provide upper bound approximations for cost efficiency measures, and that imposing convexity can improve (i.e. lower) these upper bound estimates. Indeed, convexity does not interfere with economic efficiency results and imposing it does even enhance the upper bound interpretation of technical efficiency measures in terms of economic efficiency. However, it is worth to emphasize that imposing convexity does interfere with technical (or DEA-type) efficiency analysis as such; see for instance the results for x_5.

1.4.2 Profit efficiency

For profit efficiency we obtain as dual problem for (1.18)

$$\Pi_S^+\left(z_k, \mathbb{R}_+^q\right) = \left[\max_{\kappa \in \mathbb{R}, \lambda_s \in \mathbb{R}_+} \left\{\kappa \middle| \; \textstyle\sum_{s \in S} \lambda_s z_s \geq \kappa z_k \text{ and } \textstyle\sum_{s \in S} \lambda_s = 1 \; \right\}\right]^{-1}. \quad (1.24)$$

This measure captures (the inverse of) the maximum equiproportionate *expansion* of netputs (or *scale augmentation*) within $CM\left(S\right)$. This is the McFadden gauge function as computed with respect to $CM\left(S\right)$.

$$\Pi_S^+\left(z_k, \mathbb{R}_+^q\right) = \left[\max_{\kappa \in \mathbb{R}} \left\{\kappa \middle| \; \kappa z_k \in CM(S) \; \right\}\right]^{-1}. \quad (1.25)$$

Similarly, the dual problem of (1.19) is

$$\begin{aligned} \Pi_S^-\left(z_k, \mathbb{R}_+^q\right) &= \min_{\kappa \in \mathbb{R}, \lambda_z \in \mathbb{R}_+} \left\{\kappa \middle| \sum_{s \in S} \lambda_s z_s \geq \kappa z_k \text{ and } \sum_{s \in S} \lambda_s = 1 \right\} && (1.26) \\ &= \min_{\kappa \in \mathbb{R}} \left\{\kappa \left| \kappa z_k \in CM(S) \right.\right\}. && (1.27) \end{aligned}$$

This measure captures the maximum equiproportionate netput *reduction* (or *scale reduction*) within $CM\left(S\right)$. As such, it can be labeled the 'inverse' McFadden gauge function. Expressions (1.24) and (1.26) are consistent with the established fact that imposing monotonicity and convexity on production possibilities does not affect profit efficiency analysis, i.e. $\Pi_S^I\left(z_k, \mathbb{R}_+^q\right) = \Pi_{CM(S)}^I\left(z_k, \mathbb{R}_+^q\right)$.[13]

Note that (1.24) and (1.26) reveal alternative directions of measurement to evaluate profit efficiency under incomplete price information. Both directions fit within the general directional distance function framework to evaluate (shadow) profit efficiency discussed in Färe and Grosskopf (1995) and Chambers et al. (1998). Interestingly, the benefit of the doubt idea (underlying the shadow price approach that is followed) suggests (*endogenous*) selection of

[13] See Varian (1984) and Banker and Maindiratta (1988) for more discussion.

the *most favorable* direction of measurement.[14]

This benefit of the doubt idea also gives the economic intuition behind (1.24) and (1.26). First, for any price vector under which actual (and maximum) profit is positive, the maximum netput scale expansion (within $CM(S)$) gives the minimum proportional profit expansion (compare with (1.24)). Similarly, if actual (and maximum) profit is negative, then reducing netput scale to a certain degree (within $CM(S)$) always reduces the profit loss to the same degree (compare with (1.26)).[15] Since we do not know the actual prices, we need to consider both scenarios, and the benefit-of-the-doubt idea suggests selecting the most favorable scenario (see (1.21)).

We again illustrate our discussion on profit efficiency by means of a figure. Figure 1.4 represents a situation with 4 producers that use a single input to produce a single output, i.e. $S_1' = \{1, 2, 3, 4\}$ and $z_s = (-x_s, y_s) \in \mathbb{R}_- \times \mathbb{R}_+$. Suppose that the relative input-output price corresponds to the slope of the bold iso-profit line. Under these input and output prices, the vector z_3 is profit maximizing. Note that the associated profit level is positive for these prices. The conclusion on profit efficiency clearly does not change when imposing monotonicity or convexity on the technology set. Similarly, convexity and monotonicity do not impact the degree of profit efficiency, which corresponds to the ratio of the actual profit to the maximum profit. For example, for the vector z_4, the associated profit efficiency level equals $\frac{0z_4}{0z_4'}$.

Next, we illustrate that DEA measures provide upper bound approximations for profit efficiency measures. In a situation of incomplete price information, we need to consider the possibility of positive as well as negative profit. We therefore use both the McFadden Gauge function and the 'inverse' McFadden Gauge function, as computed with respect to the convexified and monotonized technology set. For example, let us consider the shadow profit efficiency of z_4. First, for any price vector under which the profit is positive, the maximum proportionate netput scale expansion $\frac{0z_4''}{0z_4}$ gives the minimum proportional profit expansion. Second, if the profit is negative, the maximum proportionate netput reduction $\frac{0z_4'''}{0z_4}$ indicates the minimum reduction in profit loss. The shadow profit efficiency then equals

[14] See also Cherchye et al. (2010b) for an elaborated discussion of this interpretation of the Mc Fadden gauge function in terms of profit efficiency.

[15] Observe that the benefit of the doubt principle calls for selecting prices that yield actual and maximal profit with the same sign, as this guarantees the profit efficiency measure to be non-negative.

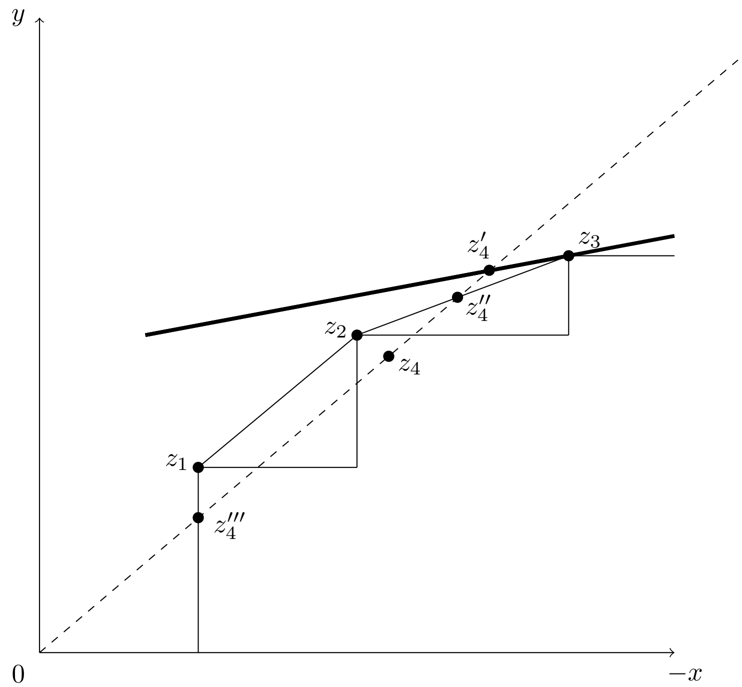


Figure 1.4: The economic meaning of DEA: profit efficiency

the maximum of $\frac{0z_4}{0z''_4}$ and $\frac{0z'''_4}{0z_4}$. For this example, the scenario of positive profit is clearly the most favorable, implying that the shadow profit efficiency equals $\frac{0z_4}{0z''_4}$ for z_4. This illustrates the upper bound interpretation of shadow profit efficiency, since $\frac{0z_4}{0z''_4} > \frac{0z_4}{0z'_4}$.

1.5 Application

In the previous sections, we highlighted the interpretation of DEA models as measuring violations of economically optimizing behavior. Our following application on efficiency in academia illustrates how crucial the specification of an appropriate behavioral model is. In a first step, we evaluate universities' efficiency in terms of cost minimization, by following the shadow price approach that we presented above. In a following step, we analyze the (possibly distortive) effects of using nonverifiable assumptions regarding behavioral objectives and technical production possibilities.

Data and variables. Our application uses data on 133 US universities in 2011-2012.[16] The universities in our sample are all reported in the top 500 of the Shanghai ranking in the year 2012. This ensures a certain degree of homogeneity in the sample. The sample consists of both public and private universities, where 89 universities are public and 44 are private. From the private universities, we only selected non-profit institutions, to guarantee that all universities in our sample share similar objectives.

We consider two inputs and three outputs. First, we use two categories of staff as university inputs: academic staff, which have teaching duties and conduct research, and other (non-academic) staff, which includes administrative and technical staff. Next, we use undergraduate student enrollments, graduate student enrollments and number of doctor's degrees granted as outputs. Table 1.1 provides some summary statistics for our selection of inputs and outputs.[17]

		Mean.	Std.	Min.	Max
Input	Academic staff	2284	1275	355	6297
	Other staff	5165	3537	963	21557
Output	Undergraduate enrollments	17833	10183	650	55016
	Graduate enrollments	5248	3216	757	16373
	Doctor's degrees	322	215	2	892

Table 1.1: Summary statistics data

Core model. We believe that cost minimization provides a suitable behavioral model to evaluate the efficiency of the universities in our sample. Indeed, we may reasonably argue that outputs are to a large extent exogenous (i.e. beyond the control of university managers), so that cost minimization for given outputs is a plausible assumption. Thus, following our argument to impose minimal prior structure (and to avoid unverifiable assumptions) for the production behavior that is evaluated, the cost efficiency measure $C^I_{OM(S)}$ (or, equivalently, its dual specification $C^I_{CIM(S)}$) constitutes an appropriate efficiency measure for the setting that we consider here. Therefore, we will use this measure as our core efficiency measure in our following analysis.

[16] We retrieved our data from the National Center for Education Statistics (NCES), which collects and analyzes detailed information on education in the United States.

[17] Staff and student enrollments are expressed in full-time equivalents.

Table 1.2 presents some summary statistics on the efficiency scores for our core model. We find that the average cost efficiency in the sample equals 0.86, which indicates that the average cost reduction potential amounts to 14%. Interestingly, the average efficiency is rather high, which conforms to our expectation that most universities effectively do operate in a (nearly) cost minimizing manner.

Next, when distinguishing between public and private universities, we observe that public universities operate on average more cost efficient than their private counterparts. Moreover, when focusing on the median efficiency measures, we find that more than 50% of the public universities turns out to be cost efficient, while the opposite conclusion applies to private universities. In this respect, a Wilcoxon rank-sum test for the null hypothesis that public and private universities achieve the same efficiency level returns a p-value of only $9,9\%$. This significant difference seems to indicate that private universities must have (slightly) different objectives than simply cost minimization.

		Mean	Min.	1^{st} Qu.	Median	3^{rd} Qu.	Max.
$C^I_{OM(S)} = C^I_{CIM(S)}$		0.86	0.28	0.76	0.98	1.00	1.00
	$\rightarrow$ public	0.90	0.33	0.81	1.00	1.00	1.00
	$\rightarrow$ private	0.83	0.30	0.68	0.97	1.00	1.00
Π^I_S		0.62	0.23	0.42	0.58	0.85	1.00
$BCC^I = C^I_{CM(S)}$		0.70	0.20	0.55	0.69	0.87	1.00

Table 1.2: Summary statistics efficiency results

Alternative specifications. In what follows, we discuss two alternative efficiency measures that build on stronger prior assumptions regarding the objectives and production possibilities underlying the observed university behavior. As we will show, these measures do indicate a considerable amount of inefficient behavior. Because the assumptions underlying the two efficiency measures are difficult to motivate a priori, we argue that the differences with the efficiency results for our core model should be interpreted as indicating an ill-specified behavioral model (and not inefficiency per se). This clearly illustrates the basic argument of our paper: erroneous production assumptions may substantially distort the efficiency analysis (and the conclusions that are drawn from it).

Our first alternative measure uses the assumption that universities pursue profit maximization instead of cost minimization. Table 1.2 gives the corresponding results for the profit efficiency measure Π_S^I. For our sample, we find that the average value of Π_S^I amounts to only 0.62, which is considerably below the average cost efficiency value that we obtained before (i.e. 0.86). This result should actually not be surprising as profit maximization is a substantially stronger assumption than cost minimization. However, we find it hard to motivate the assumption of profit maximization for the current setting (which contains only public and private non-profit institutions). Therefore, in our interpretation the substantial amount of profit inefficiency primarily reveals an ill-specified model of university behavior. In turn, this provides an additional motivation for using our core model of cost minimization.

Our second alternative measure maintains cost minimization as the behavioral objective but puts additional structure on the production possibilities. We recall that, in its dual interpretation, our core model corresponds to the measure $C_{CIM(S)}^I$, which uses the Axioms 1.1, 1.2 and 1.4. To define our alternative efficiency measure, we replace Axiom 1.4 (i.e. convexity in input space) by the stronger Axiom 1.3 (i.e. convexity in netput space). Following a similar argument as above, we can show that this obtains the shadow cost efficiency measure

$$C_{CM(S)}^I\left(z_k, \mathbb{R}_+^l\right) = \min_{\kappa \in \mathbb{R}_+, \lambda_s \in \mathbb{R}_{++}} \{\kappa | \sum_{s \in S} \lambda_s x_s \leq \kappa x_k, \sum_{s \in S} \lambda_s = 1 \text{ and } \sum_{s \in S} \lambda_s y_s \geq y_k\}. \tag{1.28}$$

Actually, the expression on the right hand side defines the technical input efficiency measure that was presented by Banker, Charnes and Cooper (1984), which is widely used in DEA applications. We denote the measure as BCC^I in Table 1.2. We find that the average value for this measure amounts to no more than 0.70, which again is significantly below the one for our core model. Similar to before, we conclude that the convexity assumption has a substantial impact on the efficiency results. In our opinion, because there is no a priori argument to use Axiom 3 as a valid production assumption for the setting at hand, this provides an additional motivation for using our core model here.

1.6 Summary and concluding discussion

We have reconsidered the economic motivation of DEA by highlighting its behavioral interpretation. Duality relationships can justify the use of certain production postulates in order to draw inference about economic efficiency performance, and so rationalize the use of certain DEA models. This potential use of DEA is all the more attractive since its engineering motivation is often unpersuasive. Importantly, the appropriate DEA model depends on the economic efficiency concept that is under consideration. In fact, this perspective may institute original efficiency evaluation models; see e.g. Cherchye et al. (2008, 2013, 2014c) who develop new nonparametric methodology for analyzing multi-output production by adopting a similar behavioral perspective of DEA.

We plead for carefully checking the validity of axioms that can interfere with the test results (e.g. Axiom 1.5 in the context of cost efficiency analysis), and for investigating the sensitivity of the results with respect to these axioms if it is difficult to verify them empirically. In our opinion, such practice falls in line with the non-parametric philosophy, which advocates minimal risk of specification error.

We also conducted an empirical application in which we demonstrated the importance of carefully specifying the behavioral production model in order to obtain a meaningful efficiency analysis. In this application we first motivated cost minimization as a good behavioral assumption to evaluate the efficiency of US universities. Next, we showed that our efficiency conclusions were very sensitive to alternative assumptions regarding the behavioral objective (profit maximization versus cost minimization) and production possibilities (convexity in netput space instead of input space).

Three further points pertain to our specification of the production-decision problem. First, the economic efficiency tests and measures discussed above implicitly assume that prices do not vary with quantities and that the eventual quantities and prices are perfectly anticipated by producers. The presented economic efficiency measures can be employed to quantify violations of these hypotheses. However, when different assumptions seem more appropriate, then the behavioral model is to be adapted, which in turn can motivate alternative DEA models (e.g. the monotone hull model (see (3)); compare with Cherchye et al.

(2000) and Kuosmanen and Post (2002)).

Second, for expositional convenience we have restricted attention to producers that seek to minimize cost or maximize profit given the production technology and the input-output prices. However, depending on the setting, alternative behavioral objectives might also be plausible. In situations in which the inputs are exogenously determined, DMUs may seek to produce the output that maximizes revenues. Moreover, even if DMUs have control over the inputs, DMUs may choose to maximize revenues instead of profits, for example to build market share and a reputation. We refer to Shepard (1970), who discusses the dual link between the revenue function and the output distance function.

Furthermore, in multi-output settings, the individual output-divisions might not always behave cooperatively. It might be the case that each individual output-division seeks to maximize the budget assigned. The budget-maximizing model was initiated in public choice theory by Niskanen (1971). In particular, the budget maximizing model states that bureaucrats seek to increase their budget in order to increase their own power. In Chapter 2, we further discuss the allocation of budget among individual output divisions and possible efficiency gains by reallocating the budget. Similarly, some individual divisions might be free riding on other divisions with respect to the use of joint inputs. We refer to Cherchye et al. (2014c), who compare the empirical validity of cooperative and noncooperative models for describing the observed production behavior.

Third, in many environments, we need to impose additional restrictions, e.g. due to the non-discretionary nature of exogenously fixed inputs or outputs or because producers face additional cost or revenue constraints (e.g. Färe et al. (1994)). In a similar vein, there may be specific characteristics of the data that should be accounted for when designing the appropriate efficiency evaluation model: input or output values may not be continuous but subject to discreteness, interval or categorical restrictions. Once more, different specifications of the production-decision problems entail alternative efficiency analysis (DEA) models.

The core idea of this chapter is that starting from a careful specification of the production-decision problem, which depends on the specific application setting, can provide economic motivation of alternative and perhaps even novel DEA models. We believe that it is important to strongly hold on to this economic perspective in practical applications, rather than 'blindly' resorting to standard, so-called 'well-established' models. In our opinion this forms a natural precondition for meaningful DEA applications.

Chapter 2

Coordination efficiency in multi-output settings: a DEA approach

Abstract

We extend a recently developed methodology for measuring the efficiency of Decision Making Units (DMUs) in the case of multiple inputs and outputs. The methodology accounts for economies of scope through the use of joint inputs, and explicitly includes information about the allocation of inputs to particular outputs. We focus on possible efficiency gains by reallocating inputs across outputs. We introduce a measure of coordination efficiency, which captures these efficiency gains. We demonstrate the practical usefulness of our methodology through an efficiency analysis of education and research conducted at US universities.[1]

[1] This chapter is based on joint work with Laurens Cherchye (KU Leuven) and Bram De Rock (ULB). I refer to the paper of Cherchye et al. (2014a), which has been published in Annals of Operations Research. We are grateful to Jo Van Biesebroeck, Frederic Vermeulen, the seminar audience in Leuven, and participants of the European Workshop on Efficiency and Productivity Analysis 2013 in Helsinki for insightful discussion. Moreover, we are grateful to two anonymous referees for their constructive comments, which improved this paper significantly.

2.1 Introduction

Data Envelopment Analysis (DEA) is a widely used approach to evaluate the efficiency of Decision Making Units (DMUs). In particular, DEA evaluates a DMU's efficiency by comparing its input and output quantities to those of other DMUs operating in a similar technological environment.[2] An attractive feature of DEA is that it is intrinsically nonparametric: DEA efficiency evaluations need not assume a specific functional/parametric form for the production technology. Instead, the production possibility set is reconstructed on the basis of the observed input-output combinations, using standard production axioms. A DMU's efficiency is then measured as the distance from its input-output combination to the frontier of this production possibility set.

Traditional DEA methods typically treat DMUs' production processes as a black box: they only use information on the aggregate amounts of inputs and outputs, and not on how the inputs and outputs are exactly linked to each other. Nevertheless, information on the allocation of inputs to outputs is often available in empirical research settings. Including this information can substantially increase the discriminatory power of the efficiency analysis, i.e. it creates considerably more potential to identify inefficient production behavior. Cherchye et al. (2013) have put this idea into practice by developing a novel DEA-based methodology for measuring the efficiency of DMUs characterized by multiple inputs and outputs. Their methodology accounts for joint inputs in the production process, and explicitly includes information on how inputs are allocated to outputs.[3]

This chapter takes this multi-output methodology one step further. We propose an extension that quantifies possible efficiency gains by reallocating inputs over outputs. We capture these efficiency gains by a new measure of coordination efficiency. The measure takes a value of one when the input allocation over outputs is efficient, while a value below unity reveals that the productive efficiency can further increase by reallocating the inputs more optimally. Interestingly, our method also provides concrete guidelines on how to achieve the better input allocation, which is especially attractive from a practical point of view.

[2]See, for example, Färe et al. (1994), Cooper et al. (2007), Fried et al. (2008) and Cook and Seiford (2009) for extensive reviews of DEA.

[3]The treatment of multi-output production is partly inspired on recent work regarding the modeling of multi-person household consumption. See Cherchye et al. (2007, 2011a,b).

We also show the empirical usefulness of our methodology through an application that evaluates the efficiency of US universities. We believe that university performance forms an interesting application area because universities typically have a two-fold assignment, i.e. education and research. In our application, we consider a university as a DMU that consists of an education division and a research division. Given our specific methodological contribution, a primal focus will be on the (efficient) allocation of university budgets over education and research outputs.

The remainder of the chapter is organized as follows. Section 2.2 motivates our research question in more detail and relates it to the relevant literature. Section 2.3 formally introduces our measure of coordination efficiency. As we will explain, this measure essentially captures the difference between so-called centralized and decentralized efficiency. Section 2.4 discusses the practical implementation of our theoretical efficiency measures. Section 2.5 shows that our distinction between centralized and decentralized efficiency also bears an interesting dual representation. Section 2.6 presents our empirical application to US universities. Finally, Section 2.7 concludes.

2.2 Multi-output production and input allocation

To set the stage, we first provide a verbal explanation of the ideas that we formalize in the following sections. Next, we also discuss the relationship between our approach and alternative approaches that have appeared in the DEA literature.

2.2.1 Centralized, decentralized and coordination efficiency

Multi-output production is often motivated by the presence of economies of scope, which originate from joint use of inputs (Cherchye et al. (2008)). Economies of scope occur if the average production cost decreases when the number of outputs increases (Baumol et al. (1982)). Scope economies typically originate from jointly (or "publicly") used inputs, i.e. inputs that simultaneously benefit the production of multiple outputs. The DEA-based method of Cherchye et al. (2013) accounts for economies of scope by explicitly modeling the presence of joint inputs in the production process.

We assume that a DMU consists of several divisions, where each division is responsible for the production of one or more outputs. To be clear, we assume that there is no overlap in the outputs that divisions produce, each output can be produced by only one particular division.[4] On the input side, we distinguish joint inputs from division-specific inputs. Division-specific inputs differ from joint inputs in that they can be allocated to the outputs produced by a particular division.[5] As we will formalize in the next section, we assume that each division is characterized by its own production technology, while accounting for interdependencies between the different technologies through joint inputs. Note that this setting corresponds to the concept 'almost non jointness', defined by Kohli (1985).[6]

As mentioned before, we are particularly interested in the way that a DMU allocates the available inputs among the divisions. More precisely, we consider whether a reallocation of the division-specific inputs within a DMU can lead to efficiency gains. To examine these gains from reallocation, we distinguish between centralized and decentralized efficiency measurement.

Essentially, the measure proposed by Cherchye et al. (2013) considers efficiency from a decentralized perspective. In particular, the allocation of the division-specific inputs is considered to be predetermined and taken for granted in the efficiency analysis. Therefore, we will refer to Cherchye et al's original measure as decentralized efficiency in the sequel.

Our following analysis will complement this measure of decentralized efficiency with an alternative (novel) measure of centralized efficiency. Intuitively, this measure assumes that a DMU's central management can reallocate inputs over output divisions. Clearly, such reallocation can give rise to new gains of productive efficiency. We quantify these additional

[4]Each DMU does not necessarily have to consist of exactly the same number of divisions. Suppose we observe M different kinds of output divisions for a particular setting. Then each DMU may consist of a subset of these divisions. Note that to be able to apply the proposed methodology, we do need to have a sample of DMUs which can be subdivided in divisions in a consistent manner, such that divisions of the same kind, produce comparable outputs for each DMU.

[5]Cherchye et al. (2013) assume that a DMU is organized in such a way that each division is responsible for just one output. In that case, the notion output-specific input is used, to be distinguished from joint input. However, perfect information about the allocation of inputs to each individual output is often not available. We therefore consider a division structure in this chapter. A main advantage of such a division structure is that it suffices to have data on the amount of inputs each division uses for the production of its outputs. Cherchye et al. (2014c) adopted a similar division structure in a setting that is formally close to the one that we consider here.

[6]The concept 'almost non-jointness' is a generalization of the concepts 'non-jointness in input' quantities (all inputs are division-specific) and 'non-jointness in prices' (all inputs are joint). See for example Samuelson (1966), Lau (1972), Hall (1973), Kohli (1983) and van den Heuvel (1986) who discuss and motivate these concepts.

efficiency gains by our measure of coordination efficiency, which we calculate as the ratio of centralized over decentralized efficiency. Note that the decentralized versus centralized setting can be linked to the standard text book argument of short run versus long run. In the short run, the division structure and the allocation of inputs might be fixed, however, in the long run the allocation of the inputs becomes variable.

In Sections 2.3 and 2.4, we will introduce our measures of centralized and decentralized efficiency as measures of input technical efficiency. Interestingly, our distinction between centralized and decentralized efficiency also has an intuitive dual interpretation in terms of cost efficiency. In dual terms, the distinction relates to the (shadow) input prices that are used to evaluate a DMU's cost efficiency, which are defined differently in the decentralized and centralized cases. This will be explained more in detail in Section 2.5.

2.2.2 Related literature

At this point, it is worth indicating that there is a close link between the approach that we develop here and other approaches that have been presented in the DEA literature. Most notably, our set-up bears direct connections to earlier work on network DEA and centralized DEA models. In a sense, our method is situated on the intersection of these two existing approaches, by combining elements that are specific to each of them.

Firstly, there is clear relation with network DEA. Network models also add additional structure to the transformation process from inputs to outputs in the DEA assessment (see Färe and Grosskopf (2000), Färe et al. (2007) and Cook and Zhu (2014)).[7] Moreover, network DEA can be used to analyze the allocation of resources across various uses. For example, Färe et al. (1997) consider the use of land to produce corn, wheat and soybeans.

Our approach has in common with network DEA that it explicitly incorporates information about the allocation of inputs to specific outputs. However, a crucial difference pertains to our dealing with joint inputs in the production model, which is not considered in the existing network DEA models. The presence of joint inputs will translate into special

[7] The model of Cherchye et al. (2013) is close in spirit to the network DEA models with parallel subsystems. However, network models can be applied to a variety of situations. For example, network DEA can also explicitly model intermediate products, which are produced and used within the production process. Similarly, a network formulation might be a dynamic model in which some outputs in period t are used as an input in period $t + 1$.

constraints that link the production process of the outputs, constraints that are absent in conventional network DEA models. We refer to Lozano (2015), who discusses in detail how the approach of Cherchye et al. (2013) can be combined with existing network models.

As indicated above, by including joint inputs in the analysis, we effectively model the presence of economies of scope, which forms a prime economic motivation for simultaneously producing multiple outputs. With respect to economies of scope, it is worth to mention the studies of Tone and Sahoo (2003) and Sahoo and Tone (2013), who model scale and scope economies of firms due to process indivisibilities arising from the task-specific production processes of a multi-product firm.

Secondly, Lozano and Villa (2004) introduced so-called centralized DEA models. These models assume that there is a centralized decision maker who "owns" or supervises the DMUs. The centralized decision maker is interested in maximizing the efficiency of each individual unit, but is also concerned about the total input consumption. Lozano and Villa (2004) assume that inputs can be reallocated across DMUs and seek for an optimal allocation of the inputs among DMUs.

Following Lozano and Villa (2004), we also use the term "centralized" efficiency for a setting where input reallocation is possible. However, a main distinguishing feature of our approach is that we allocate inputs among various uses *within* the same DMU, i.e. across alternative production technologies (associated with different output divisions). This contrasts with centralized DEA models, which reallocate inputs across DMUs that are characterized by identical production technologies.

In this chapter, we assume that each division produces different outputs. However, in a setting with divisions that are not responsible for the production of different outputs, but rather produce a similar set of outputs, the centralized efficiency model, introduced by Lozano and Villa (2004), would be more appropriate to study a reallocation.

2.3 Theoretical efficiency measures

After introducing some necessary notation and terminology, we will formally define our measures of centralized, decentralized and coordination efficiency. To structure our discussion, we will first consider a theoretical set-up in which the production technology is known. The next section will consider the practical implementation of our theoretical efficiency measures.

2.3.1 Preliminaries

Efficiency analysis starts from a data set with T observed DMUs that produce N outputs. We assume that each DMU is subdivided into M divisions, where each division m ($1 \leq m \leq M$) produces one or more outputs. Let d^m represent the number of outputs produced by division m. Thus, we have $N = d^1 + \ldots + d^M$. For each DMU t ($1 \leq t \leq T$), we observe the vector of produced outputs $\mathbf{y}_t \in \mathbb{R}^N_+$, where

$$\mathbf{y}_t = (\mathbf{y}_t^1, \ldots, \mathbf{y}_t^M), \tag{2.1}$$

so that $\mathbf{y}_t^m \in \mathbb{R}^{d^m}_+$ denotes the vector of outputs produced by division m. Similarly, we also observe the division-specific and joint inputs. For any output division m, the vector $\mathbf{q}_t^m \in \mathbb{R}^{Nspec}_+$ contains the division-specific inputs. We let $\mathbf{q}_t = \sum_{m=1}^M \mathbf{q}_t^m$, i.e. $\mathbf{q}_t$ represents the total division-specific inputs of DMU t. Finally, $\mathbf{Q}_t \in \mathbb{R}^{Njoin}_+$ represents the joint inputs of DMU t. We note that these joint inputs $\mathbf{Q}_t$ cannot be allocated to particular divisions; they are simultaneously used in the production process of all output divisions. Taken together, the empirical analysis starts from the following data set:

$$S = \{(\mathbf{y}_t, \mathbf{q}_t^1, \ldots, \mathbf{q}_t^M, \mathbf{Q}_t) \mid t = 1, \ldots, T\}. \tag{2.2}$$

Next, we consider a separate production technology for each output division. Importantly, we account for interdependencies between the different technologies through jointly used inputs. More formally, we characterize the production technology of a division m by input requirement sets $I^m(\mathbf{y}^m)$, which contain all combinations of division-specific and joint inputs $(\mathbf{q}^m, \mathbf{Q})$ that can produce the output quantities $\mathbf{y}^m$, i.e.

$$I^m(\mathbf{y}^m) = \{(\mathbf{q}^m, \mathbf{Q}) \in \mathbb{R}^{Nspec}_+ \times \mathbb{R}^{Njoin}_+ | (\mathbf{q}^m, \mathbf{Q}) \text{ can produce } \mathbf{y}^m\}. \tag{2.3}$$

Finally, as explained above, our centralized efficiency measure assumes that the central management of a DMU coordinates the production process of all divisions. More specifically, the central management determines how much of the total amount $\mathbf{q}$ of division-specific inputs goes to every division. To formalize this idea, we need to consider input requirement sets $\mathbf{I}(\mathbf{y})$ for the "aggregate" output vector $\mathbf{y}$ (defined over all divisions simultaneously). These input requirement sets are constructed from the division-specific sets $I^m(\mathbf{y}^m)$, as follows:

$$\mathbf{I}(\mathbf{y}) = \{(\mathbf{q}, \mathbf{Q}) \mid \exists\, \mathbf{q}^1, \ldots, \mathbf{q}^M \text{ such that } \sum_{m=1}^{M} \mathbf{q}^m = \mathbf{q} \text{ and } \forall m : (\mathbf{q}^m, \mathbf{Q}) \in I^m(\mathbf{y}^m)\}. \tag{2.4}$$

Thus, each set $\mathbf{I}(\mathbf{y})$ contains all combinations of division-specific and joint inputs $(\mathbf{q}, \mathbf{Q})$ that can produce the output $\mathbf{y} = (\mathbf{y}^1, \ldots, \mathbf{y}^M)$. In particular, we say that $(\mathbf{q}, \mathbf{Q})$ can produce $\mathbf{y}$ if the input $\mathbf{q}$ can be allocated among the divisions such that every division can produce the associated output $\mathbf{y}^m$.

2.3.2 Efficiency measures

Throughout, we will consider input-oriented efficiency measurement, which identifies the maximum possible input reduction while keeping the output fixed. In doing so, we adopt radial (or Debreu-Farrell) efficiency measures, which are most popular in applied DEA work. Essentially, for some evaluated DMU and a given technology (represented by input requirement sets), these radial measures seek the maximum equiproportionate input reduction for the given output. Attractively, the measures have a natural degree interpretation: they are situated between 0 and 1, with an efficiency score of unity indicating technically efficient production, while lower values reflect greater productive inefficiency. Finally, and importantly, radial measures also have an interesting dual representation in terms of cost efficiency, which we will illustrate in Section 2.5.

Let us first introduce our radial measure of decentralized efficiency TE_t^d, which -to recall- is the one that was originally proposed by Cherchye et al. (2013). In formal terms, we have

$$TE_t^d = \min\{\theta \mid \forall m : (\theta \mathbf{q}_t^m, \theta \mathbf{Q}_t) \in I^m(\mathbf{y}_t^m)\}. \tag{2.5}$$

This measure captures the maximum equiproportionate reduction of the inputs (captured by θ) that is feasible, conditional on the given (observed) allocation of division-specific inputs among the divisions.

By contrast, our (novel) centralized efficiency measure TE_t^c no longer takes the observed input allocation for granted. It quantifies the maximum input reduction that is feasible while accounting for possible (optimal) reallocation of the division-specific inputs. We define

$$TE_t^c = \min\{\theta \mid (\theta \mathbf{q}_t, \theta \mathbf{Q}_t) \in \mathbf{I}(\mathbf{y}_t)\}. \tag{2.6}$$

Basically, the measure considers not only the production process of the individual divisions (like the measure TE_t^d) *but also* the coordination among divisions. In other words, in contrast to our measure of decentralized efficiency, our centralized efficiency measure takes into account the inefficiencies that result from a suboptimal allocation of the inputs to the output divisions.

This directly provides the basic intuition of our following result, which states that decentralized efficiency is never lower than centralized efficiency.[8]

Proposition 2.1. We have that $TE^c \leq TE^d$.

In turn, this motivates the following ratio measure C^oE_t as a natural measure for the efficiency of input coordination among divisions:

$$C^oE_t = \frac{TE_t^c}{TE_t^d}. \tag{2.7}$$

By construction, this coordination efficiency measure is situated between 0 and 1. A coordination efficiency value of unity indicates that the inputs are allocated in an optimal way across the divisions. By contrast, a lower efficiency value reveals that DMU t's productive efficiency can be further increased by input reallocation. More specifically, for the given output DMU t can achieve additional input reduction by adjusting the input mix over output divisions.

[8] The proofs of our results appear in Appendix 2.B.

As a final note, we can also write

$$TE_t^c = C^o E_t \times TE_t^d, \tag{2.8}$$

which provides an intuitive decomposition of centralized efficiency as the product of coordination efficiency and decentralized efficiency.[9]

2.4 Practical implementation

In practice, the true input requirement sets $I^m(\mathbf{y}^m)$ (used for the decentralized measure TE_t^d) and $\mathbf{I}(\mathbf{y})$ (used for the centralized measure TE_t^c) are typically not observed. The DEA approach proceeds by defining empirical approximations of these input sets on the basis of some standard production axioms. In turn, this defines operational efficiency measures that can be computed by means of standard linear programming techniques.

2.4.1 Empirical efficiency measures

In what follows, we use the same production axioms as Cherchye et al. (2013):

Axiom 2.1 (Nested input sets)**.** $\mathbf{y}^m \geq \mathbf{y}^{m*} \Rightarrow I^m(\mathbf{y}^m) \subset I^m(\mathbf{y}^{m*})$

Axiom 2.2 (Monotone input sets)**.** $(\mathbf{q}^m, \mathbf{Q}) \in I^m(\mathbf{y}^m)$ and $(\mathbf{q}^{m*}, \mathbf{Q}^*) \geq (\mathbf{q}^m, \mathbf{Q}) \Rightarrow (\mathbf{q}^{m*}, \mathbf{Q}^*) \in I^m(\mathbf{y}^m)$

Axiom 2.3 (Convex input sets)**.** $(\mathbf{q}^m, \mathbf{Q}), (\mathbf{q}^{m*}, \mathbf{Q}^*) \in I^m(\mathbf{y}^m) \Rightarrow$
$\forall \lambda \in [0,1] : \lambda(\mathbf{q}^m, \mathbf{Q}) + (1-\lambda)(\mathbf{q}^{m*}, \mathbf{Q}^*) \in I^m(\mathbf{y}^m)$

Axiom 2.4 (Observability means feasibility)**.** $(\mathbf{y}_t, \mathbf{q}_t^1, \ldots, \mathbf{q}_t^M, \mathbf{Q}_t) \in S \Rightarrow$
$\forall m : (\mathbf{q}_t^m, \mathbf{Q}_t) \in I^m(\mathbf{y}_t^m)$.

In words, Axiom 2.1 says that, if some input can produce the output $\mathbf{y}^m$, then it can also produce any lower output $\mathbf{y}^{m*}$. Essentially, this means that outputs are freely disposable.

[9] Note that it is possible to consider coordination efficiency in a dynamic context, and further decompose the productivity index. See Grosskopf (2003) for a discussion on the Malmquist productivity index and its decomposition. Moreover, Ang and Kerstens (2016) introduce a nonparametric measure of coordination productivity growth, based on the Luenberger productivity indicator. The coordination productivity indicator is decomposed into a coordination technical inefficiency change component and a coordination technical change component.

Similarly, Axiom 2.2 defines free input disposability, i.e. more input never reduces the output. Next, Axiom 2.3 states that, if two inputs can produce the output $\mathbf{y}^m$, then any convex combination of these inputs can also produce the same output. Finally, Axiom 2.4 says that the observed input-output combinations are certainly feasible.[10]

Of course, other axioms can be used as well. For example, such axioms may impose alternative assumptions regarding the nature of returns-to-scale (constant, decreasing or increasing) that underlie the production technology. For compactness, we do not extensively discuss such assumptions here. However, it is worth emphasizing that these assumptions can be incorporated into our method, by combining our method with the work of Petersen (1990) and Bogetoft (1996). An interesting feature of our methodology is that it allows for imposing returns-to-scale assumptions that are specific to individual divisions.[11]

Cherchye et al. (2013) have shown that, if Axioms 2.1-2.4 hold, an empirical inner bound approximation of the true (but unobserved) set $I^m(\mathbf{y}_t^m)$ is defined as

$$\hat{I}(\mathbf{y}_t^m) = \{(\mathbf{q}^m, \mathbf{Q}) | \sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}^m, \sum_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \mathbf{Q}, \sum_{s \in D_t^m} \lambda_s^m = 1, \lambda_s^m \geq 0\}, \tag{2.9}$$

with

$$D_t^m = \{s | \mathbf{y}_t^m \leq \mathbf{y}_s^m\} \tag{2.10}$$

the set of dominating DMUs.[12] By construction, we have $\hat{I}^m(\mathbf{y}_t^m) \subseteq I^m(\mathbf{y}_t^m)$ under Axioms 2.1-2.4. The set $\hat{I}(\mathbf{y}_t^m)$ is the smallest production set that is consistent with Axioms 2.1-2.4 for a given data set S and, therefore, it defines a useful empirical approximation of the true set $I^m(\mathbf{y}_t^m)$.

[10] Essentially, this assumes that the input and output data are not contaminated by measurement errors. The DEA literature has forwarded alternative proposals to deal with errors-in-the-data in practical applications. To compactify our exposition, we will abstract from measurement issues in what follows, but these existing methodologies are fairly easily integrated in the framework that we set out here. See Cherchye et al. (2013) for related discussion.

[11] The resulting production technologies can then also be used to estimate division specific returns-to-scale; see Podinovski (2004c) for more discussion.

[12] Note that for each division m, D_t^m contains by construction only one element for the division with the highest output level. As such, these divisions are efficient by default. This is a well-known issue in DEA and is due to the absence of making assumptions regarding to the returns-to-scale in the production process. A solution is to work with a constant returns-to-scale technology. Under the assumption of constant returns-to-scale, an increase in inputs results in a proportionate increase in the output level. As a result, both small and large divisions can be compared with each other.

By combining (2.4) and (2.9), we obtain the following empirical approximation of the set $\widehat{\mathbf{I}}(\mathbf{y})$:

$$\widehat{\mathbf{I}}(\mathbf{y}) = \{(\mathbf{q}, \mathbf{Q}) \mid \exists\, \mathbf{q}^1, \ldots, \mathbf{q}^M \text{ such that } \sum_m \mathbf{q}^m = \mathbf{q} \text{ and } \forall m : (\mathbf{q}^m, Q) \in \hat{I}(\mathbf{y}^m)\}. \tag{2.11}$$

Clearly, because $\hat{I}^m(\mathbf{y}_t^m) \subseteq I^m(\mathbf{y}_t^m)$ we also have that $\hat{\mathbf{I}}(\mathbf{y}_t) \subseteq \mathbf{I}(\mathbf{y}_t)$.

Using the empirical approximations $\hat{I}^m(\mathbf{y}_t^m)$ and $\hat{\mathbf{I}}(\mathbf{y}_t)$, we can define the empirical counterparts of TE_t^d and TE_t^c as, respectively,

$$\widehat{TE_t^d} = \min\{\theta | \forall m : (\theta \mathbf{q}_t^m, \theta \mathbf{Q}_t) \in \hat{I}^m(\mathbf{y}_t^m)\}, \tag{2.12}$$

and

$$\widehat{TE_t^c} = \min\{\theta \mid (\theta \mathbf{q}_t, \theta \mathbf{Q}_t) \in \widehat{\mathbf{I}}(\mathbf{y}_t)\}, \tag{2.13}$$

which have a readily analogous interpretation as the theoretical measures. Interestingly, the empirical measures $\widehat{TE_t^d}$ and $\widehat{TE_t^c}$ can be computed by solving simple linear programming problems, which we discuss in more detail below.

Before doing so, we point out two properties of the measures that are relevant for our following exposition. Firstly, because $\hat{I}^m(\mathbf{y}_t^m) \subseteq I^m(\mathbf{y}_t^m)$ and $\hat{\mathbf{I}}(\mathbf{y}_t) \subseteq \mathbf{I}(\mathbf{y}_t)$, we naturally obtain

$$TE_t^d \leq \widehat{TE_t^d} \quad \text{and} \quad TE_t^c \leq \widehat{TE_t^c}, \tag{2.14}$$

i.e. the empirical efficiency measures define natural upper bounds for the theoretical measures. Secondly, just as in the theoretical case, the centralized efficiency measure $\widehat{TE_t^c}$ never exceeds the decentralized measure $\widehat{TE_t^d}$.

Proposition 2.2. We have that $\widehat{TE^c} \leq \widehat{TE^d}$.

Analogous to before, the result in Proposition 2.2 motivates the empirical measure for coordination efficiency

$$\widehat{C^oE_t} = \frac{\widehat{TE_t^c}}{\widehat{TE_t^d}}, \tag{2.15}$$

which has a similar meaning as the theoretical measure C^oE_t. Again, a low coordination efficiency reveals that efficiency gains are possible by reallocating the inputs in a more optimal way.

2.4.2 Linear programming formulation

We conclude this section by discussing the linear programming formulation of $\widehat{TE}_t^d$ and $\widehat{TE}_t^c$. By using (2.9) and (2.11), it is straightforward to verify that

$$
\begin{array}{lrcl}
\textbf{(LP-1)} & \widehat{TE}_t^d & = & \min\limits_{\theta_t \geq 0, \lambda_s^m \geq 0} \theta_t \\
& s.t. & & \\
& \text{(D-1)} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \theta_t \mathbf{Q}_t \\
& \text{(D-2)} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \theta_t \mathbf{q}_t^m \\
& \text{(D-3)} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m = 1;
\end{array}
$$

$$
\begin{array}{lrcl}
\textbf{(LP-2)} & \widehat{TE}_t^c & = & \min\limits_{\theta_t \geq 0, \lambda_s^m \geq 0, \mathbf{q}^m \geq 0} \theta_t \\
& s.t. & & \\
& \text{(D-0)} & & \sum\limits_{m} \mathbf{q}^m = \theta_t \mathbf{q}_t \\
& \text{(D-1)} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \theta_t \mathbf{Q}_t \\
& \text{(D-2)'} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}^m \\
& \text{(D-3)} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m = 1.
\end{array}
$$

Both **(LP-1)** and **(LP-2)** compute efficiency measures that quantify the maximum possible input reduction (captured by θ_t) for DMU t, when t is compared to its dominating DMUs.[13] Before solving the linear programs, we need to identify the set D_t^m of dominating

[13] Note that in these LPs we do not use convex combinations of the outputs to define possible comparison partners

DMUs for each division. To be concrete, for each division m, we need to find the DMUs for which $\mathbf{y}_s^m \geq \mathbf{y}_t^m$. Subsequently, the linear program compares for each division m, the inputs for DMU t to convex combinations of the corresponding inputs for the dominating DMUs. In particular, for each division m a reference vector is constructed on the basis of a convex combination of all DMUs s that produce at least the output $\mathbf{y}_t^m$. Each variable λ_s^m then represents the weight of every DMU s in this convex combination. These variables are also called "intensity parameters" in the DEA literature.

The essential difference between **(LP-1)** and **(LP-2)** pertains to the variables $\mathbf{q}^m$ that appear in **(LP-2)**. Specifically, in **(LP-2)** the input quantities $\mathbf{q}^m$ can be chosen freely (except from the non-negativity requirement and the adding-up constraint (D-0), i.e. $\sum_m \mathbf{q}^m = \theta_t \mathbf{q}_t$), whereas in **(LP-1)** the division-specific inputs are fixed at their observed level $\mathbf{q}_t^m$. This additional freedom to choose the quantities $\mathbf{q}^m$ in **(LP-2)** also directly explains the inequality $\widehat{TE^c} \leq \widehat{TE^d}$ in Proposition 2.2.

Interestingly, the linear programming problem **(LP-2)** not only defines an empirical measure of centralized efficiency. It also returns reference values for the division-specific inputs (i.e. the solution values for $\mathbf{q}^m$) that correspond to an optimal input allocation. This is especially attractive from a practical point of view. For DMUs that have a low coordination efficiency, it provides specific guidelines on how to improve the input allocation over output divisions.

One final note pertains to the possibility that input reallocations may be restricted in practice. For example, it may well be that some inputs are simply not adjustable and/or that changing the input mix is very expensive. Conveniently, it is fairly easy to adapt our framework to take such considerations into account. In particular, as long as the associated restrictions can be formulated in linear form, they can simply be added to **(LP-2)** without interfering with the linear nature of the resulting programming problem. For compactness, and because we believe this type of extension is relatively straightforward, we will not further elaborate on this.

for the given DMU. That is, we only consider DMUs that effectively produce more output. This follows from the fact that in Axiom 3 we do not assume convexity in both inputs and outputs simultaneously. As discussed in depth in Cherchye and Post (2003), this assumption is often problematic in practical applications. However, it is fairly straightforward to change Axiom 3, and correspondingly the linear programs **(LP-1)** and **(LP-2)**, to allow for convexity in both inputs and outputs.

2.5 Dual representations

An interesting feature of our measures of centralized and decentralized efficiency is that they can be given a dual representation as measures of cost efficiency, evaluated at shadow input prices. The difference between the two measures relates to the input prices that are used for evaluating the division-specific inputs. As we will explain, this difference can be given a precise interpretation in terms of centralized versus decentralized decision making (i.e. with versus without input reallocations across output divisions).

2.5.1 Decentralized efficiency

We first consider our measure of decentralized efficiency. The dual version of the linear program **(LP-1)** can be written as[14]

$$
\begin{aligned}
\textbf{(LP-3)} \qquad \widehat{TE}_t^d \quad &= \max_{\substack{c_t^m \geq 0, \mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}, \\ \mathbf{P}_t \in \mathbb{R}_+^{Njoin}, \mathbf{p}_t^m \in \mathbb{R}_+^{Nspec}}} \sum_{m=1}^{M} c_t^m \\
s.t.& \\
\text{(C-1)} \qquad & \sum_{m=1}^{M} \mathbf{P}_t^m = \mathbf{P}_t \\
\text{(C-2)} \qquad & \forall m : c_t^m \leq (\mathbf{p}_t^m)' \, \mathbf{q}_s^m + (\mathbf{P}_t^m)' \, \mathbf{Q}_s \; \forall s \in D_t^m \\
\text{(C-3)} \qquad & \sum_{m=1}^{M} (\mathbf{p}_t^m)' \, \mathbf{q}_t^m + \mathbf{P}_t' \mathbf{Q}_t = 1.
\end{aligned}
$$

To explain the cost efficiency interpretation of this dual program, we need to interpret the vectors $\mathbf{P}_t^m$, $\mathbf{P}_t$ and $\mathbf{p}_t^m$ as shadow price vectors. For every division m, $\mathbf{p}_t^m$ and $\mathbf{P}_t^m$ contain the shadow prices for the division-specific and joint inputs, respectively. Similarly, $\mathbf{P}_t$ contains the prices for the joint inputs at the level of the aggregate DMU.

For the joint inputs, the division-specific prices $\mathbf{P}_t^m$ are related to the DMU-level prices $\mathbf{P}_t$ by the adding-up restriction (C-1). This adding-up restriction implies that the shadow prices $\mathbf{P}_t^m$ actually represent the fractions of DMU t's aggregate prices $\mathbf{P}_t$ that are borne

[14] Appendix 2.A gives specific details on how we obtain the dual problems **(LP-3)** and **(LP-4)**.

by each division m. In a sense, they represent the "willingness-to-pay" (reflecting marginal productivities) of the different divisions for the jointly (or "publicly") consumed inputs. This parallels the interpretation of so-called Lindahl prices that correspond to the efficient provision of public goods.

Next, the constraint (C-3) defines a cost normalization for the evaluated DMU t. It specifies that the shadow prices must be such that the aggregate cost of DMU t equals unity.

Given all this, $\sum_{m=1}^{M} c_t^m$ can be interpreted as the minimum cost for producing the output of DMU t. Each variable c_t^m then represents the minimum cost for producing the output of DMU t's division m, while accounting for the interrelation with the output production of the other divisions (through joint inputs). In particular, restriction (C-2) imposes that c_t^m cannot exceed the cost level associated with any other DMU s that produces at least the output $\mathbf{y}_t^m$ (i.e. $s \in D_t^m$).

In this respect, we also note that, by construction, $t \in D_t^m$ for any m. Therefore, the normalization constraint (C-3) guarantees that

$$\sum_{m=1}^{M} c_t^m \leq 1 (= \sum_{m=1}^{M} (\mathbf{p}_t^m)' \, \mathbf{q}_t^m + \mathbf{P}_t' \mathbf{Q}_t)), \tag{2.16}$$

which implies that the sum of division-specific minimal costs $\sum_{m=1}^{M} c_t^m$ is situated between 0 (because of the non-negativity constraints) and 1. Conveniently, because DMU t's aggregate cost level is normalized at unity, this also makes that the sum $\sum_{m=1}^{M} c_t^m$ in the objective of problem **(LP-3)** can be interpreted as the ratio of minimal cost (for the aggregate output $\mathbf{y}_t$) over DMU t's actual cost. Putting it differently, the objective function value of **(LP-3)** expresses DMU t's cost efficiency in relative terms.

The max operator in the objective guarantees that the shadow prices $\mathbf{p}_t^m$, $\mathbf{P}_t^m$ and $\mathbf{P}_t$ are chosen such that this measure of cost efficiency is maximized. In a sense, this actually gives the "benefit-of-the-doubt" to the evaluated DMU. Most favorable prices are chosen, so putting DMU t in the best possible light.

One important final note is in order. In program **(LP-3)** the prices $\mathbf{p}_t^m$ for the division-specific inputs may vary depending on the division m at hand. Intuitively, the fact that different divisions can use different (shadow) prices for these inputs reflects that these inputs are not directly substitutable across divisions. This effectively relates to the very essence of our notion of decentralized efficiency, which assumes that input reallocations across divisions are impossible. It will also imply a crucial difference with the dual representation of our centralized efficiency measure.

2.5.2 Centralized efficiency

We next turn to our centralized efficiency measure. The dual of program **(LP-2)** is given as

$$\textbf{(LP-4)} \quad \widehat{TE_t^c} = \max_{\substack{c_t^m \geq 0, \mathbf{P}_t^m \in \mathbb{R}_+^{Njoin}, \\ \mathbf{P}_t \in \mathbb{R}_+^{Njoin}, \mathbf{p}_t \in \mathbb{R}_+^{Nspec}}} \sum_{m=1}^{M} c_t^m$$

$$s.t.$$

$$\text{(C-1)} \qquad \sum_{m=1}^{M} \mathbf{P}_t^m = \mathbf{P}_t$$

$$\text{(C-2)} \qquad \forall m : c_t^m \leq \mathbf{p}_t' \mathbf{q}_s^m + \left(\mathbf{P}_t^m\right)' \mathbf{Q}_s \; \forall s \in D_t^m$$

$$\text{(C-3)} \qquad \sum_{m=1}^{M} \mathbf{p}_t' \mathbf{q}_t^m + \mathbf{P}_t' \mathbf{Q}_t = 1.$$

This linear programming problem has basically the same structure as problem **(LP-3)**. Therefore, it also has a directly similar cost efficiency interpretation. However, there is a subtle but important difference, which pertains to the shadow prices for the division-specific inputs. In the new problem **(LP-4)**, these prices are the same for all divisions m (i.e. $\mathbf{p}_t^m$ = $\mathbf{p}_t$ for all m). The intuition directly relates to our concept of centralized efficiency: in contrast to the decentralized efficiency setting, we now assume that input reallocations over divisions are possible, which means that division-specific inputs are perfectly substitutable across divisions (and, thus, we use common division-specific shadow prices).

A last remark relates to the possibility to impose restrictions on input substitutability across divisions. As discussed at the end of Section 2.4, restrictions on input reallocations could be implemented as linear constraints added to the primal problem **(LP-2)**. In a similar vein, we could also include additional restrictions to the dual problem **(LP-4)**. It follows from our above discussion that particular constraints on the (non)substitutability of inputs can here be implemented in the form of shadow price restrictions. For example, one may want to impose that some division-specific inputs are perfectly substitutable across divisions (resulting in common shadow prices for all divisions) while other inputs are not (implying division-specific shadow prices).

2.6 Application: US university education and research

We illustrate the practical usefulness of our methodology through an empirical application that evaluates the efficiency of US universities. In particular, we are interested in the allocation of the university budget across education and research divisions. We measure inputs as expenditures, which we subdivide into division-specific and joint expenses. In this application, division-specific inputs are university expenses that are clearly directed towards either education or research. Next, joint inputs contain expenditures related to "public" services like libraries, museums, media, technology and administration. In what follows, we will first motivate our input and output data in more detail, and subsequently present our main efficiency results.

2.6.1 Input and output data

When it comes to evaluating university efficiency, the definition of the relevant input and output dimensions is all but straightforward. There is a lack of consensus in the literature on the most appropriate selection of inputs and outputs.[15] To focus our discussion, and given that our objective here is mainly to illustrate the practical application of our methodology, we opt for a fairly basic specification of a university's production process.

We are interested in the joint production of education and research conducted by universities, and we will particularly concentrate on the allocation of the university budget across

[15] In particular, there is a lack of consensus in the literature whether research funding should be modeled as an input or as an output of the universities. De Witte and López-Torres (2015) provide an extensive overview of studies that model research funding either as an input or as an output. The difficulty to categorize research funding either as an input or an output might be the result of an endogeneity problem. In particular, the ability of the universities to attract research funding depends on the quality of the research output they produce. Mayston (2015) argues that a research grant should therefore be seen as an endogenous resource input. We refer to Mayston (2015) and references therein for a discussion and suggestions to handle endogeneity within a DEA framework. To keep the exposition simple, we chose to model research income as an input in the university production process, however, one should keep the endogeneity of research income in mind when interpreting the results.

the education and research divisions. In this respect, our study here is close in spirit to Beasley (1995), who analyzed the teaching and research efficiency of chemistry and physics departments in the UK. This author also takes into account that some resources are shared between the different activities. However, an important difference is that Beasley (1995) considered the shared input as a division-specific input of which the allocation over research and teaching is unknown. By contrast, our specific methodology allows us to treat shared inputs as joint inputs that simultaneously benefit the production of both teaching and research. We must note that we therefore assume in this chapter that economies of scope between teaching and research are present. In principle, this assumption can also be tested, see for example De Witte et al. (2013), who investigate economies of scope in research and teaching. The authors find a (limited) presence of economies of scope. Moreover, De Witte et al. (2013) provide an interesting literature overview on the debate of the precise relationship between teaching and research.

We use data on 130 US universities in 2012. Our sample contains both (87) public and (43) private non-profit institutions. We retrieved the main part of our data from the National Center for Education Statistics (NCES). The NCES collects and analyzes detailed information on education in the United States. In order to obtain a sample that is sufficiently comparable (i.e. homogeneous), we (only) selected universities that are classified as research universities, with high or very high research activity (according to the Carnegie classification). Moreover, our selected universities all appeared in the top 500 of the Shanghai ranking in the year 2012.[16]

Output selection. We consider two outputs for the education division and two outputs for the research division of each university. The two education outputs are undergraduate and graduate enrollments, expressed in full time equivalents. The advantage of considering enrollments (and not degrees granted) is that this measure also takes into account students who have not yet completed their studies but did receive education from the university. Our enrollment data come from the NCES and pertain to the academic year 2011-2012.

The two research outputs are number of doctor's degrees and a measure of publication

[16]The Shanghai ranking is a widely used world ranking of universities that simultaneously accounts for several indicators of research performance. See www.shanghairanking.com for more details.

output. Again, number of doctor's degrees is taken from the NCES and relates to the year 2012. Next, our measure of publication output is constructed such that it not only accounts for the quantity but also the quality of scientific publications, where quality is measured in terms of researchers' citations. Specifically, we quantify publication output as the mean of 3 scores that are also used for the Shanghai ranking: HiCi (i.e. highly cited researchers in 21 broad subject categories), N&S (i.e. papers published in Nature and Science) and PUB (i.e. papers indexed in Science Citation Index-expanded and Social Science Citation Index). Although the total Shanghai score is computed as a weighted sum of 6 individual criteria, we have selected the three individual criteria which are directly related to publications. We take the mean of those three criteria, to construct a proxy for publications. In this respect, we follow the Shanghai ranking, which also gives equal weight to these three criteria to compute the total Shanghai score. These three scores are standardized between 0 and 100. In particular, the highest scoring institution is assigned a score of 100, and the scores of the other institutions are calculated as a percentage of the top score.

We would like to make two further remarks with respect to our measures of research output. First, in our analysis we start from the standardized Shanghai scores, to construct a proxy for publication output. However, if one would have access to the raw data, one could construct a composite indicator, see Nardo et al. (2008) for a handbook on constructing composite indicators. In principle, when construction a composite indicator, one needs to go through a number of steps, which involve modeling choices to be made in each step. In this respect, an interesting reference is Cherchye et al (2008), who discuss creating composite indicators with DEA and robustness analysis with respect to these modeling choices.

Second, we have opted for a fairly basis specification of the research output. However, there exist many alternative options to quantify research performance. For example, Vinkler (1998), Vinkler (2006), Ruiz et al. (2010) and Bonaccorsi et al. (2006) have developed multidimensional indicators for evaluating the research performances of research institutes. Alternatively, De Witte and Rogge (2010) propose a methodology to aggregate multi-dimensional research output, to evaluate individual researcher performances.

Input selection. We use university expenses as inputs. The public universities report expenses according to GASB (Governmental Accounting Standards Board), while the private universities report according to FASB (Financial Accounting Standards Board).[17] We distinguish between expenses that can be allocated directly to the research and eduction divisions and expenses that have a joint (or "public") nature.[18]

Our division-specific inputs can be retrieved directly from the NCES data. Specifically, we use as education input all expenses that relate to activities that form part of the instruction program. Similarly, our research input contains all expenses related to activities specifically organized to produce research outcomes. As well the instruction expenses as the research expenses consist for a substantial amount of salaries and wages. At this point it is worth to note that many faculty members combine instruction and research. To allocate the salaries and wages of those faculty members, one would ideally have data on the percentage of time each faculty member devotes to instruction and to research. Since such detailed information is not available, a second best option is to distinguish between instructional faculty and research faculty on the basis of their primary function. Specifically, the NCES asks the universities to classify instructional faculty as those members of the instruction/research staff whose major regular assignment is instruction, including those with released time for research. Instructional faculty also include faculty for whom it is not possible to differentiate between teaching, research and public service, because each of these functions is an integral component of his/her regular assignment. Research faculty are those members of the instruction/research staff whose primary function is research. The universities report the instructional and research expenses based on this criterion. Note that we need to keep this in mind when interpreting the coordination efficiency results.

Finally, we note that for these division-specific inputs we work with expenses (so wages times staff), which avoids the discussion of the quality of staff in the following sense. Uni-

[17] Our analysis implicitly assumes that GASB and FASB are comparable accounting standards. In this respect, we also conducted several robustness checks to ensure that our conclusions with respect to public and private universities cannot be attributed to differences in these accounting standards. In particular, we computed efficiency results with numbers of staff (instead of expenses) as inputs. Interestingly, these alternative exercises yielded the same qualitative conclusions.

[18] In principle, alternative scenarios are possible. We could consider particular expenses as joint instead of division-specific and vice versa. Given the illustrative nature of the application we have chosen not to do so for compactness.

versities with better qualified faculty will in general offer higher salaries and wages, which in turn leads to more expenses. As such our efficiency analysis will take this quality difference (implicitly) into account. See, e.g., Camanho and Dyson (2005, 2008), Fukuyama and Weber (2008) and Sahoo et al. (2014) for recent discussion on the relevance of taking the quality of the inputs into account in a DEA analysis.

Next, our joint input contains expenses on institutional and academic support, which are vital to the proper "overall" functioning of a university (including both education and research). Institutional support includes, for example, the management, personnel administration and logistic activities. Academic support contains academic administration, libraries, museums and computer services.

We remark that our following analysis will not consider expenses on public services (e.g. community service programs, radio, television and consulting) or student services (e.g. student activities, newspapers, health services and athletics). Our motivation is that we believe these expenses are not directly related to the education and research outputs that we selected. However, it should be clear that these data could easily be included in the analysis if deemed appropriate. Our choice not to use this information is purely an empirical one and does by no means indicate a limitation of our methodology.

Data. Appendix 2.C contains our input and output data for each individual university. Appendix 2.C also reports the number of universities in the dominating sets, for both instruction and research. Table 2.1 reports some summary statistics. Expenses are reported in millions of dollars. We find substantial variation across our sample of universities for each of the variables that we selected. See, for example, the high standard deviations and the large differences between minimum and maximum values.

The last two columns of Table 2.1 report the mean values for the public and private universities. A main difference between these two categories of universities relates to the number of undergraduate enrollments. On average, public universities have 22551 undergraduate enrollments, which is more than twice the average of 9177 enrollments for the private universities. In this respect, a remarkable observation is that average expenses on instruction for the public universities are nearly one third below those for their private counterparts.

Variables	Full sample				Public	Private
	Mean	Std.	Min.	Max.	Mean	Mean
Instruction and research expenses	766	560	142	2774	690	921
→ Instruction	458	343	88	1947	399	579
→ Research	308	259	4	1291	291	342
Academic and institutional exp.	242	186	50	1434	205	318
Undergraduate enrollment	18127	10065	968	55016	22551	9177
Graduate enrollment	5318	3189	757	16373	5271	5413
Doctor's degree	329	213	35	892	350	284
Publication	30	16	11	100	28	35

Table 2.1: Descriptive statistics for the sample of 130 US universities, including 87 public and 43 private universities. Expenses are reported in millions of dollars.

On the basis of these figures, public universities seem to be more efficient than private universities in the provision of (undergraduate) education. However, an important remark is that student enrollments only measure the quantity but not the quality of this education output. For example, it may well be that private universities specialize in "excellent" education, while public universities rather focus on "standard" education. Unfortunately, we could not find data on the educational quality for our sample of universities, and so we are bound to ignore such quality considerations in our following efficiency analysis. While we will not repeat it explicitly, this qualification must be kept in mind when interpreting our efficiency results.

2.6.2 Efficiency results

We will first report summary statistics for our measures of decentralized, centralized and coordination efficiency, defined over our sample of 130 universities. Subsequently, we will consider efficiency differences between public and private universities. In a final step, we will take a closer look at the allocation of university budgets across education and research. In particular, starting from the observation that our sample is characterized by substantial coordination inefficiency, we use the results of our linear programs to identify possible strategies that can lead universities towards a more optimal input allocation.

Full sample results. Table 2.2 provides summary results for our different efficiency measures. Detailed information on the efficiency results of each university can be found in Appendix 2.C.

If we first consider centralized efficiency for the full sample of universities, we find that about 40% of the universities is labeled as efficient. The average centralized efficiency amounts to 0.86, which suggests that the average university can reduce its expenses by 14% for the given levels of education and research outputs.

As indicated above, centralized efficiency can be decomposed into decentralized efficiency and coordination efficiency. We find that the average decentralized efficiency equals 0.92. This indicates that the possible input reduction amounts to only 8% if input reallocations over research and education are impossible.

The difference between centralized and decentralized efficiency is captured by our measure of coordination efficiency. The average coordination efficiency turns out to be 0.93, which reveals that optimal input reallocations can yield an additional efficiency gain of 7%. Interestingly, because the median coordination efficiency value is 0.98 (i.e. below unity), we conclude that such input reallocations can be beneficial to more than half of the universities in our sample. [19]

		Min.	1^{st} Qu.	Mean	Median	3^{rd} Qu.	Max.
Centralized Eff.		0.40	0.76	0.86	0.92	1.00	1.00
	$\rightarrow$ public	0.45	0.80	0.89	0.97	1.00	1.00
	$\rightarrow$ private	0.40	0.66	0.80	0.85	0.99	1.00
Decentralized Eff.		0.50	0.88	0.92	1.00	1.00	1.00
	$\rightarrow$ public	0.57	0.90	0.94	1.00	1.00	1.00
	$\rightarrow$ private	0.50	0.75	0.88	1.00	1.00	1.00
Coordination Eff.		0.62	0.90	0.93	0.98	1.00	1.00
	$\rightarrow$ public	0.64	0.91	0.94	1.00	1.00	1.00
	$\rightarrow$ private	0.62	0.84	0.91	0.94	1.00	1.00

Table 2.2: Descriptive statistics for our efficiency results

[19] As a robustness exercise, we have added the expenses on student services to the instruction expenses. The average expenses on student services amount to 69 million dollar per year. Although this is a substantial category of expenses for the universities, this category is relatively small compared to the instruction expenses, which is on average 458 million dollar. Including the expenses on student service has little impact on the decentralized, centralized and coordination efficiency scores. Remarkably, the average efficiency scores remain the same. Still, the efficiency scores of the individual universities might vary a little bit.

Public versus private universities. When distinguishing between the public and private universities, we observe that the public universities operate on average more efficient than their private counterparts. This difference is more pronounced for our measures of decentralized and centralized efficiency than for our measure of coordination efficiency. This suggests that the better performance of public universities is not so much the result of a better input allocation per se (i.e. coordination efficiency), but rather follows from a more efficient input use for the given allocation of expenses over education and research (i.e. decentralized efficiency).

We also conducted Wilcoxon rank-sum tests to evaluate the statistical significance of these observed efficiency differences. We find that the difference is significant for the centralized efficiency measure (the null hypothesis that both subsamples achieve the same efficiency level has a p-value of only 0.004) and the decentralized efficiency measure (p-value of 0.024). By contrast, there is no significant difference in terms of coordination efficiency (p-value of 0.109).

Generally, we may conclude that we find a rather substantial difference in centralized efficiency between private and public universities, and decentralized efficiency seems to be a more important explanation of this difference than coordination efficiency. Our outcome that public universities perform better than private universities falls in line with the results in Chapter 1, where we performed a similar analysis without imposing additional structure on the production of education and research. Moreover, our outcome conforms with earlier findings of Ahn et al. (1988), who compared the relative efficiencies of public and private doctoral-granting universities in the US. These authors equally concluded that public universities prove to be more efficient than private universities when managerial and program inefficiencies are present in the data. However, it is worth to note there is no consensus in the literature about the impact of public or private ownership on the efficiency of schools in general (De Witte and López-Torres (2015)). It is sometimes argued that the level of student performance would be higher in private institutions. However, studies taking into account student performance find mixed results. For example, Duncombe et al. (1997) find that efficiency of school districts is negatively associated with the relative number of private school students. Similarly, Agasisti (2013) find that private schools perform worse than public ones. On the other hand, Cherchye et al. (2010a) and Ferrera et al. (2011) find that the type of ownership does not seem to account for differences in student performance.

Reallocation strategies. Our above analysis indicates that only 49 of the 130 universities achieve a centralized efficiency score of one. The remaining 81 universities have a score below one, which means that efficiency improvements are possible. When zooming in on these 81 universities, we find that 73 of them exhibit coordination inefficiency. For these universities, reallocating the division-specific inputs over the education and research divisions effectively leads to efficiency gains.

In this respect, we recall that a centralized efficiency score of θ indicates that all expenses (i.e. joint expenses as well as division-specific expenses) can be reduced by a fraction $(1-\theta)$. However, and importantly, these savings need not be equally distributed across the education and research divisions. It is even possible that one of the divisions receives additional (division-specific) budget as a consequence of the reallocation, at the expense of the other division.

As explained above (when discussing **(LP-2)**), our linear programs not only define an empirical measure of centralized efficiency, but also provide specific guidelines on how to enhance the input allocation over output divisions. Specifically, it returns reference values for the division-specific inputs that correspond to an optimal input allocation. Interestingly, this information enables us to conclude whether a division needs to reduce or increase the current (division-specific) input, in order to remedy the observed coordination inefficiency.

In Appendix 2.C, we report for each university the changes in the budget directed towards education or research that are necessary to eliminate its coordination inefficiency. Figure 2.1 provides a schematic summary of the different strategies that can be followed, hereby also indicating to how many universities each strategy applies. In the first scenario, both divisions need to reduce their inputs, but not in equal proportion. In percentage terms, one of the divisions needs to save considerably more than the other division. This scenario holds for 38 universities. For 20 universities, the education division needs to reduce the division-specific input more than the research division. For the remaining 18 universities, the opposite conclusion holds.

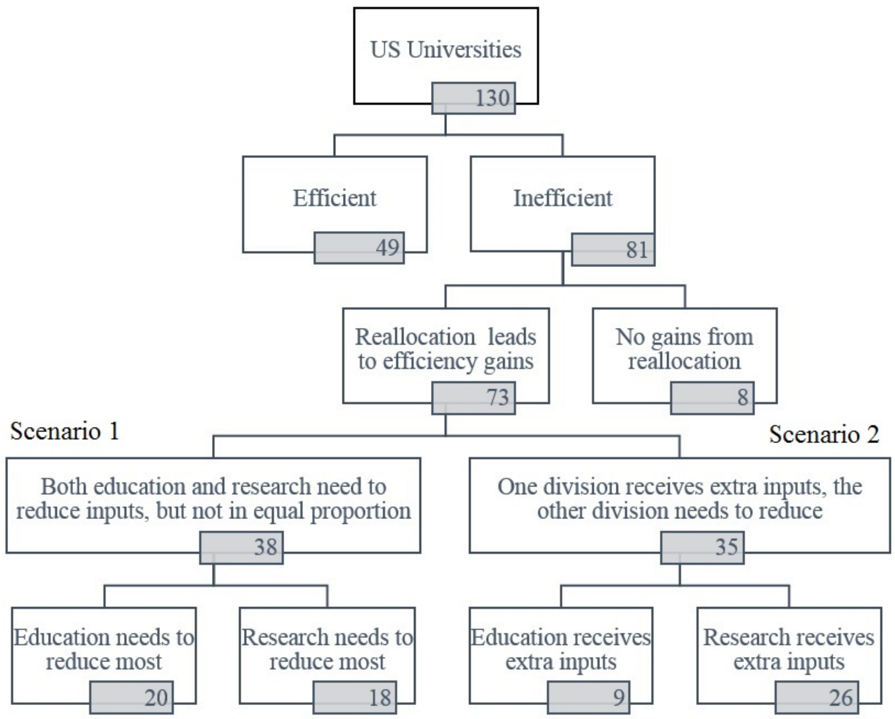


Figure 2.1: Possible strategies to achieve an optimal allocation of the budget.

In the second scenario, which holds for 35 universities, one of the divisions receives additional division-specific input. Of course, in such a case the required input reduction for the other division is particularly substantial, since the input use at the aggregate DMU level (summed over the two divisions) needs to go down. For this scenario, we find that the research division should receive extra budget (at the expense of the education division) in 26 of the 35 universities, while the opposite conclusion applies to the other 9 universities.

2.6.3 Shadow prices

Finally, we come back to the dual interpretation of the centralized and decentralized efficiency measures. As mentioned before, the dual interpretation in terms of cost efficiency assigns a (division-specific) shadow price to the inputs. In the application on university education and research, we have expenses figuring as inputs. We denote the instruction expenses with q^I, the research expenses with q^R and the academic and institutional expenses with Q. These categories of expenses have division-specific shadow prices, with I indicating the instruction division and R indicating the research division. In particular, the instruction and research expenses, which figure as division-specific inputs, have shadow prices p^I and p^R and the academic and institutional expenses, which figure as joint input, have shadow price $P = P^I + P^R$. Note that in the centralized model, the shadow prices p^I and p^R of the instruction and research expenses are assumed to be equal.

Tables 2.3 and 2.4 report the values of the shadow prices, computed for the decentralized and centralized model. The reported shadow prices correspond to expenses which are expressed in billions of dollars.[20] We refer to table 2.8 in Appendix 2.C for the shadow prices and corresponding weights for every university.

Furthermore, we also report w^I and w^R, the share of the total budget that is allocated to the instruction division and the research division, when computing the cost efficiency of the universities. Specifically, we define

$$w^I = \frac{p^I q^I + P^I Q}{p^I q^I + p^R q^R + PQ},$$

[20] Note that the magnitude of the shadow prices depends on the scaling of the expenditure categories. However, the relative shadow prices are independent of the scale and indicate the relative weight of one expenditure category over the other, when cost efficiency is computed.

		Min.	1^{st} Qu.	Mean	Median	3^{rd} Qu.	Max.
	p^I	0	0	2,08	0	3,68	16,67
Shadow	p^R	0	0	4,28	0	4,42	52,03
Prices	P^I	0	0	2,12	0	0	33,51
	P^R	0	0	3,62	0	3,49	38,95
Budget	w^I	0	0	0,41	0	1	1
Shares	w^R	0	0	0,59	1	1	1

Table 2.3: Summary shadow prices and budget shares for decentralized model

		Min.	1^{st} Qu.	Mean	Median	3^{rd} Qu.	Max.
Shadow	p	0	0	1,56	0,63	2,46	11,11
Prices	P^I	0	0	3,56	0	4,08	38,34
	P^R	0	0	4,80	1,38	6,03	38,95
Budget	w^I	0	0	0,49	0,53	0,80	1,00
shares	w^R	0	0,2	0,51	0,47	1,00	1,00

Table 2.4: Summary shadow prices and budget shares for centralized model

$$w^R = \frac{p^R q^R + P^R Q}{p^I q^I + p^R q^R + PQ}.$$

We find that for the decentralized model, the share of expenditures allocated to the instruction output is always either 0 or 1 and vice versa for the share allocated to the research output. This is the result of many shadow prices having a value equal to zero. The fact that the benefit of the doubt measurement assigns all weight to only one of the two objectives of the universities might be a very unrealistic scenario. For the centralized model, there are more universities that have shadow prices different form zero and the share of the total budget that is allocated to the instruction and research division, is more balanced. Still, there are many shadow prices estimated to be zero.

Shadow price restrictions It is a well-recognized problem that many DEA applications obtain unrealistic shadow prices. We refer to Cook and Seiford (2009) for a nice overview of the literature that addresses the problem of unacceptable shadow prices. These methodologies usually impose shadow price restrictions, either in a direct or an indirect way. Such shadow price restrictions may rule out extreme cases where the relative price of the inputs approaches zero or infinity.

We will impose shadow price restrictions to make sure that the 'benefit of the doubt' budget share of each expenditure category (computed with the shadow prices), does not deviate too much from the real budget share. In particular, we impose that the benefit of the doubt budget share of q^I, q^R and Q has to be maximum twice as large and minimum half as large as the real budget share:

$$\frac{1}{2}\frac{q^I}{q^I+q^R+Q} \leq \frac{p^I q^I}{p^I q^I + p^R q^R + PQ} \leq 2\frac{q^I}{q^I+q^R+Q}$$

$$\frac{1}{2}\frac{q^R}{q^I+q^R+Q} \leq \frac{p^R q^R}{p^I q^I + p^R q^R + PQ} \leq 2\frac{q^R}{q^I+q^R+Q}$$

$$\frac{1}{2}\frac{Q}{q^I+q^R+Q} \leq \frac{PQ}{p^I q^I + p^R q^R + PQ} \leq 2\frac{Q}{q^I+q^R+Q}.$$

Note that we do not impose any restrictions on the division-specific price of the joint expenses. In particular, it remains possible that the joint expenditures are fully borne by one of the two divisions.

		Min.	1^{st} Qu.	Mean	Median	3^{rd} Qu.	Max.
	TE^d	0,41	0,74	0,84	0,88	0,97	1,00
	p^I	0,30	1,01	2,74	2,09	3,64	14,76
Shadow	p^R	0,30	1,32	3,09	2,13	3,96	12,59
Prices	P^I	0	0	1,34	0,35	1,94	9,09
	P^R	0	0,18	2,41	1,20	3,41	16,92
Budget	w^I	0,13	0,28	0,50	0,47	0,73	0,99
Shares	w^R	0,01	0,27	0,50	0,53	0,72	0,87

Table 2.5: Summary shadow prices and budget shares for decentralized model, with price restrictions.

		Min.	1^{st} Qu.	Mean	Median	3^{rd} Qu.	Max.
	TE^c	0,35	0,70	0,80	0,84	0,95	1
Shadow	p	0,30	1,21	2,40	2,23	3,27	8,62
Prices	P^I	0	0	2,06	0,69	2,86	16,70
	P^R	0	0,20	2,64	1,70	3,60	14,69
Budget	w^I	0,16	0,33	0,53	0,49	0,72	0,91
shares	w^R	0,09	0,28	0,47	0,51	0,67	0,84

Table 2.6: Summary shadow prices and budget shares for centralized model, with price restrictions.

Tables 2.5 and 2.6 report the values of the shadow prices and budget shares, computed for the decentralized and centralized model, taking into account the previously mentioned shadow price restrictions. Due to these price restrictions, there are no zero shadow prices for the division-specific inputs. The division-specific shadow price of the joint inputs might still be zero. This indicates that the willingness to pay for the public inputs may be zero. The first line of the tables also reports summary statistics of the efficiency scores, taking into account the shadow price restrictions. The decentralized efficiency score is now on average equal to 0.84, in contrast to an average score of 0.92 in a setting without price restrictions. Similarly, the centralized efficiency score is now on average equal to 0.80, compared to 0.86 without price restrictions. The difference between the decentralized and centralized efficiency scores has become smaller, with the assumption of price restrictions. We refer again to Appendix 2.C for the results for all the universities. On average, the coordination efficiency is now equal to 0.96, which is slightly higher than in the setting without price restrictions.

2.7 Conclusion

We have extended the DEA-based methodology of Cherchye et al. (2013) for multi-output efficiency measurement. Our extension exploits the specific feature of this methodology, which includes information on joint and division-specific inputs in the efficiency evaluation. In particular, we use this input information to develop a method that investigates whether input reallocations across output divisions can yield specific efficiency gains. We propose a measure of coordination efficiency to quantify these gains. Interestingly, for DMUs with low coordination efficiency our method also provides concrete guidelines to achieve a more optimal input allocation. At this point it is worth to note that the proposed methodology is demanding in terms of data availability. If good data on the allocation of inputs to outputs is available, the methodology leads to a rich and powerful analysis. However, in many settings, such detailed data is still not collected. In terms of future research, more attention to the collection of detailed data would greatly advance the potential to apply multi-output efficiency measurement. An interesting way to collect data on the internal production processes is through activity based costing (ABC) systems (Cooper and Kaplan (1988)). In ABC costing systems, the costs (or inputs) are first allocated to activities and, subsequently, these activity costs are allocated to the products (or outputs).

We have used our methodology to evaluate the productive efficiency of education and research conducted at US universities. We believe our methodology is particularly well-suited to analyze the joint production of education and research, as it can account for both joint inputs and inputs specifically allocated to education or research. Although our application was mainly intended to serve illustrative purposes, it did clearly reveal the potential of our new method. For example, our empirical results suggest that the universities under study can considerably enhance their productive efficiency by adopting a more optimal input allocation (over education and research outputs). In particular, we found that more than half of the universities suffer from coordination inefficiency. Note that better data with respect to the allocation of expenses to education and research would increase the reliability of the coordination efficiency scores. However, our estimates on coordination efficiency are consistent with how the universities themselves report their expenses.

2.A Dual formulations

In this section, we clarify the link between the primal and dual formulation of the LP-models. In the decentralized setting, the linear programming problem **(LP-3)** is dual to problem **(LP-1)**. Similarly, in the centralized setting, problem **(LP-4)** is dual to problem **(LP-2)**.

We start by considering the decentralized setting. The constraint (D-1) in **(LP-1)**, which constructs for every division m a reference vector for the joint inputs, corresponds to the shadow price vector $\mathbf{P}_t^m$ in the dual cost efficiency problem. Similarly, the constraint (D-2), which constructs a benchmark vector for the division-specific inputs, corresponds to the shadow price vector $\mathbf{p}_t^m$. Further, the constraint (D-3) corresponds to the (minimal) costs variable c_t^m for every division m.

Next, we focus on the centralized setting. The only difference between **(LP-3)** and **(LP-4)** is that **(LP-4)** uses the common price vector $\mathbf{p}_t$ for the division-specific inputs. To see that **(LP-4)** is dual to the linear programming problem **(LP-2)**, we first write the problem **(LP-2)** in a slightly different form. For this, we use lemma 2.1:

Lemma 2.1. The statement

$$\exists\, \mathbf{q}^1, \ldots, \mathbf{q}^M \text{ such that } \sum_m \mathbf{q}^m = \theta_t \mathbf{q}_t \text{ and } \forall m : \sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}^m \tag{2.17}$$

is equivalent to the statement

$$\sum_m \sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \theta_t \mathbf{q}_t. \tag{2.18}$$

As a consequence, the linear programming problem **(LP-2)** is equivalent to the following problem **(LP-2b)**:

$$\begin{array}{lllll}
\textbf{(LP-2b)} & \widehat{TE}_t^c & = & \min\limits_{\theta_t \geq 0, \lambda_s^m \geq 0} \theta_t & \\
 & s.t. & & & \\
 & \text{(D-1)} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m \mathbf{Q}_s \leq \theta_t \mathbf{Q}_t & \\
 & \text{(D-2)"} & & \sum\limits_m \sum\limits_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \theta_t \mathbf{q}_t & \\
 & \text{(D-3)} & & \forall m : \sum\limits_{s \in D_t^m} \lambda_s^m = 1 &
\end{array}$$

Note that this linear programming problem **(LP-2b)** is very similar to **(LP-1)**. The only difference is that left hand side of constraint (D-2)" contains a sum over all divisions m (whereas the constraint (D-2) specifies a separate constraint for every individual m). The constraint (D-2)" corresponds to the shadow price vector $\mathbf{p}_t$ in the dual problem. We conclude that problem **(LP-4)** is dual to problem **(LP-2b)** and, thus, also to the equivalent problem **(LP-2)**.

2.B Proofs

Proposition 2.1. For $\mathbf{q}_t = \sum_m \mathbf{q}_t^m$ we have

$$\{\theta \mid \forall m : (\theta \mathbf{q}_t^m, \theta \mathbf{Q}_t) \in I^m(\mathbf{y}_t^m)\} \subset \left\{ \theta \;\middle|\; \begin{array}{c} \exists\, \mathbf{q}^1, \ldots, \mathbf{q}^M \text{ such that } \sum_m \mathbf{q}^m = \mathbf{q}_t \\ \text{and } \forall m : (\theta \mathbf{q}^m, \theta \mathbf{Q}_t) \in I^m(\mathbf{y}_t^m) \end{array} \right\}$$
$$= \{\theta \mid (\theta \mathbf{q}_t, \theta \mathbf{Q}_t) \in \mathbf{I}(\mathbf{y}_t)\}$$

Consequently,

$$\min\{\theta \mid (\theta \mathbf{q}_t, \theta \mathbf{Q}_t) \in \mathbf{I}(\mathbf{y}_t)\} \leq \min\{\theta \mid \forall m : (\theta \mathbf{q}_t^m, \theta \mathbf{Q}_t) \in I^m(\mathbf{y}_t^m)\},$$

which obtains that $TE^c \leq TE^d$. □

Proposition 2.2. We will prove that constraint (D-2) implies the constraints (D-0) and (D-2)'. Suppose that (D-2) holds. Define $\mathbf{q}^m = \theta_t \mathbf{q}_t^m$. We have found values for $\mathbf{q}^1, \ldots, \mathbf{q}^M$ such that (D-0) and (D-2)' hold. We conclude that the feasible region of **(LP-1)** is a subset of the feasible region of **(LP-2)**. Since **(LP-1)** and **(LP-2)** are both minimization problems, the optimal objective function value for **(LP-2)** cannot exceed the optimal value for **(LP-1)**. We conclude that $\widehat{TE^c} \leq \widehat{TE^d}$. □

Lemma 2.1. It is straightforward that (2.17) implies (2.18). Furthermore, suppose that (2.18) holds. Define

$$\mathbf{q}^m = \sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m,$$

for $m = 1, \ldots, M-1$, and

$$\mathbf{q}^M = \theta_t \mathbf{q}_t - \sum_{m=1}^{M-1} \mathbf{q}^m.$$

Then, by construction $\sum_m \mathbf{q}_t^m = \theta_t \mathbf{q}_t$ and $\sum_{s \in D_t^m} \lambda_s^m \mathbf{q}_s^m \leq \mathbf{q}^m$, which implies that statement (2.17) is satisfied. □

2.C Data and results

Universities	Inputs Expenses ($ million)			Outputs Enrollment (FTE)			Research	Dominating sets		Efficiency scores			Reallocation of budget for	
	Instr	Res	A&I	Undergr	Grad	Doc	Pub	$\|D^i\|$	$\|D^r\|$	TE^d	TE^c	C^oE	Instr	Res
Arizona State University	637	243	369	55016	10479	611	33,1	1	18	1,00	1,00	1,00	1,00	1,00
Auburn University	264	131	121	19777	3137	247	14	50	71	0,92	0,78	0,85	0,71	0,93
Boston College	243	37	171	9525	3573	149	14,6	69	102	1,00	0,76	0,76	0,64	1,58
Boston University	838	194	264	20951	8692	507	35,9	11	22	1,00	1,00	1,00	1,00	1,00
Brandeis University	133	47	51	3970	2644	82	15,3	108	108	1,00	1,00	1,00	1,00	1,00
Brigham Young University-Provo	424	43	145	31364	2775	92	15,5	14	106	1,00	0,87	0,87	0,78	1,72
Brown University	254	124	180	6109	1864	232	34,5	109	44	1,00	0,76	0,76	0,46	1,38
California Institute of Technology	203	281	118	968	1286	172	56,1	124	12	1,00	1,00	1,00	1,00	1,00
Carnegie Mellon University	337	221	199	5834	5048	284	27,5	57	49	0,75	0,62	0,83	0,46	0,86
Case Western Reserve University	270	395	113	4177	2469	186	26,6	110	60	0,82	0,82	1,00	0,72	0,89
Clemson University	211	144	76	16035	3297	220	16,5	65	79	1,00	1,00	1,00	1,00	1,00
Colorado State University-Fort Collins	241	218	103	21891	4014	235	26,2	36	57	1,00	1,00	1,00	1,00	1,00
Columbia University in the City of New York	1947	671	357	8163	14946	558	58,8	2	8	1,00	1,00	1,00	1,00	1,00
Cornell University	471	378	474	14887	5463	501	52,5	38	11	1,00	0,90	0,90	0,59	1,28
Dartmouth College	153	161	321	4237	1630	73	23,1	116	74	0,76	0,66	0,86	0,76	0,55
Drexel University	299	119	313	16433	5388	163	15,2	36	96	0,92	0,85	0,92	0,92	0,67
Duke University	901	899	464	8109	6159	450	50,4	35	14	0,74	0,62	0,83	0,31	0,93
Emory University	532	409	271	8495	5699	243	35,4	40	40	0,53	0,51	0,97	0,53	0,49
Florida State University	330	160	126	31800	6824	428	24,1	8	39	1,00	1,00	1,00	1,00	1,00
George Mason University	303	80	115	19594	6547	212	16	19	82	1,00	1,00	1,00	1,00	1,00
George Washington University	398	140	269	10206	8356	224	18,9	16	78	1,00	0,97	0,97	1,04	0,80
Georgetown University	416	172	288	7440	4911	116	17,1	58	98	0,50	0,41	0,82	0,37	0,50
Georgia Institute of Technology-Main Campus	282	658	137	14517	7076	483	29,8	22	28	1,00	1,00	1,00	1,00	1,00
Harvard University	1064	769	1434	9515	13315	691	100	3	1	1,00	1,00	1,00	1,00	1,00
Indiana University-Bloomington	530	113	205	32420	8514	468	29,1	5	31	1,00	1,00	1,00	1,00	1,00
Indiana University-Purdue University-Ind.	399	180	229	18794	8311	35	15,9	14	111	1,00	0,91	0,91	1,31	0,02
Iowa State University	257	195	191	24128	3203	376	27,5	33	43	1,00	1,00	1,00	1,00	1,00
Johns Hopkins University	1520	1254	438	6452	13979	479	53,9	3	11	0,92	0,85	0,93	1,14	0,50
Kansas State University	211	156	92	17593	2569	162	17,1	70	93	1,00	1,00	1,00	1,00	1,00
Kent State University at Kent	180	23	106	20660	4709	142	15,2	37	101	1,00	1,00	1,00	1,00	1,00
Lehigh University	130	34	87	5239	1227	101	14,1	115	111	1,00	1,00	1,00	1,00	1,00
Louisiana State University	274	275	150	23195	4094	322	23,1	33	53	0,83	0,83	1,00	0,83	0,83
Massachusetts Institute of Technology	643	1291	618	4364	6398	573	66,4	36	4	0,58	0,58	1,00	0,58	0,58
Michigan State University	620	380	232	35003	5912	491	34,9	5	26	0,95	0,95	1,00	0,95	0,94
Montana State University	88	127	50	11314	1039	53	13,1	90	123	1,00	1,00	1,00	1,00	1,00
New York University	1202	644	417	24402	16373	417	46,3	1	17	1,00	1,00	1,00	1,00	1,00
North Carolina State University at Raleigh	386	267	162	24394	5993	446	27,3	20	35	1,00	0,97	0,97	0,78	1,26
Northeastern University	327	91	196	17934	6653	125	16,7	22	97	0,94	0,89	0,94	0,88	0,94
Northwestern University	631	426	476	9017	8331	378	48,2	18	17	0,97	0,86	0,90	0,63	1,21

Universities	Inputs Expenses ($ million)			Outputs Enrollment (FTE)		Research		Dominating sets		Efficiency scores			Reallocation of budget for	
	Instr	Res	A&I	Undergr	Grad	Doc	Pub	$\lvert D^i \rvert$	$\lvert D^r \rvert$	TE^d	TE^c	C^oE	Instr	Res
Ohio State University-Main Campus	922	490	402	45479	10545	756	41,8	1	7	1,00	1,00	1,00	1,00	1,00
Oregon State University	219	197	119	19003	3138	197	26,7	55	58	0,84	0,78	0,93	0,93	0,62
Pennsylvania State University-Main Campus	1170	766	618	41350	5612	629	45,5	4	9	0,82	0,52	0,64	0,24	0,96
Princeton University	381	273	312	5240	2839	351	50,4	98	16	1,00	0,94	0,94	0,41	1,69
Purdue University-Main Campus	620	268	231	31592	7276	649	35,1	8	14	1,00	0,97	0,97	0,85	1,23
Rensselaer Polytechnic Institute	142	117	83	5622	1165	136	17,7	115	92	1,00	1,00	1,00	1,00	1,00
Rice University	251	89	90	3774	2614	190	25,3	110	62	1,00	1,00	1,00	1,00	1,00
Rutgers University-New Brunswick	824	342	280	32676	6246	414	36	6	29	1,00	0,87	0,87	0,72	1,25
Saint Louis University-Main Campus	235	44	107	9175	2442	205	12,9	93	86	1,00	0,87	0,87	0,66	2,01
San Diego State University	190	4	92	23874	3294	48	18	34	95	1,00	1,00	1,00	1,00	1,00
Southern Methodist University	153	22	142	6333	2759	67	10,7	100	122	1,00	1,00	1,00	1,00	1,00
Stanford University	1200	1023	549	7485	6749	764	76,8	30	1	1,00	1,00	1,00	1,00	1,00
Stony Brook University	426	123	225	16544	5515	263	25,9	35	55	0,92	0,72	0,78	0,65	0,94
SUNY at Albany	198	231	141	12475	2900	158	16,6	77	94	0,81	0,64	0,79	0,95	0,37
Syracuse University	312	73	197	16429	3776	150	15,9	58	98	1,00	0,67	0,67	0,58	1,08
Temple University	420	110	264	26598	3828	216	17,2	22	79	1,00	0,74	0,74	0,66	1,03
Texas A & M University-College Station	591	538	200	36801	7613	663	36,5	3	13	1,00	1,00	1,00	1,00	1,00
Texas Tech University	202	137	125	23897	4232	254	14,9	31	69	1,00	1,00	1,00	1,00	1,00
The University of Montana	97	45	58	11645	1196	44	14,9	89	119	1,00	1,00	1,00	1,00	1,00
The University of Tennessee	554	289	252	19668	5289	461	24,7	31	33	0,70	0,64	0,91	0,51	0,89
The University of Texas at Austin	772	549	358	35361	9049	867	45,4	3	3	1,00	1,00	1,00	1,00	1,00
The University of Texas at Dallas	155	86	82	10549	5487	181	17,1	40	88	1,00	1,00	1,00	1,00	1,00
The University of Texas at San Antonio	160	51	109	22898	3008	77	12,9	38	119	1,00	0,94	0,94	1,08	0,49
Tufts University	212	137	288	4728	4442	143	24,5	72	70	0,73	0,70	0,95	0,73	0,65
Tulane University of Louisiana	258	157	104	8257	2642	120	17,8	93	93	0,84	0,84	1,00	0,83	0,85
University at Buffalo	390	146	244	20380	4730	305	20,5	37	55	0,80	0,74	0,91	0,69	0,86
University of Alabama at Birmingham	273	268	279	9915	4661	174	25	57	66	0,57	0,45	0,79	0,57	0,33
University of Alaska Fairbanks	104	145	69	5749	757	50	15,7	117	110	1,00	1,00	1,00	1,00	1,00
University of Arizona	430	467	297	29451	5319	446	38,3	15	26	0,92	0,84	0,92	0,73	0,94
University of Arkansas	184	125	106	18154	2704	164	14,9	64	97	0,95	0,88	0,93	1,02	0,67
University of California-Berkeley	647	629	359	27737	9680	892	68,6	7	1	1,00	1,00	1,00	1,00	1,00
University of California-Davis	652	614	378	25315	4955	566	47,2	25	10	1,00	0,81	0,81	0,38	1,27
University of California-Irvine	507	317	261	23472	5013	413	37,6	29	28	1,00	0,96	0,96	0,68	1,40
University of California-Los Angeles	1451	833	670	27911	10929	725	60,5	4	3	0,76	0,54	0,71	0,41	0,76
University of California-Riverside	212	121	93	18182	2463	263	29,3	64	48	1,00	1,00	1,00	1,00	1,00
University of California-San Diego	631	829	459	24272	5306	523	59	27	8	0,78	0,78	1,00	0,78	0,78
University of California-Santa Barbara	235	203	108	19439	3392	346	38,4	49	30	1,00	1,00	1,00	1,00	1,00
University of California-Santa Cruz	144	132	87	16220	1607	172	31,4	76	49	1,00	1,00	1,00	1,00	1,00
University of Central Florida	276	115	164	45446	5852	229	19,1	3	76	1,00	1,00	1,00	1,00	1,00

Universities	Inputs Expenses ($ million)			Outputs Enrollment (FTE)		Research		Dominating sets		Efficiency scores			Reallocation of budget for	
	Instr	Res	A&I	Undergr	Grad	Doc	Pub	$\|D^i\|$	$\|D^r\|$	TE^d	TE^c	C^oE	Instr	Res
University of Chicago	990	319	322	4988	7035	401	47,3	29	16	1,00	0,83	0,83	0,29	2,52
University of Cincinnati-Main Campus	304	209	223	20875	7592	242	23,5	15	61	0,94	0,78	0,83	0,94	0,55
University of Colorado Boulder	378	321	145	25523	3423	344	41,2	25	29	1,00	1,00	1,00	1,00	1,00
University of Colorado Denver	408	291	97	9746	5364	107	21,6	44	78	0,90	0,90	1,00	0,90	0,90
University of Connecticut	477	163	337	17998	4379	313	20,1	48	56	0,69	0,49	0,71	0,38	0,82
University of Delaware	346	135	141	18067	3510	228	22	54	66	0,88	0,74	0,84	0,68	0,89
University of Florida	708	609	288	32257	11226	696	39,3	2	11	1,00	1,00	1,00	1,00	1,00
University of Georgia	286	352	173	25346	7599	453	29,7	12	31	1,00	1,00	1,00	1,00	1,00
University of Hawaii at Manoa	295	350	92	13443	2811	196	29,5	78	51	1,00	1,00	1,00	1,00	1,00
University of Houston	248	115	220	26555	5017	301	22,7	22	54	1,00	1,00	1,00	1,00	0,98
University of Illinois at Chicago	664	296	173	16024	7758	342	28,7	16	46	1,00	1,00	1,00	1,00	1,00
University of Illinois at Urbana-Champaign	589	489	323	34331	12847	869	45,4	1	2	1,00	1,00	1,00	1,00	1,00
University of Iowa	389	338	235	22609	3471	437	32,6	36	32	0,96	0,89	0,94	0,52	1,32
University of Kansas	375	298	174	18874	4787	302	20,8	38	55	0,78	0,70	0,90	0,80	0,57
University of Kentucky	291	293	210	19363	3397	322	19,5	49	56	0,62	0,59	0,95	0,62	0,56
University of Louisville	272	161	135	13668	2847	185	13,3	77	92	0,68	0,66	0,96	0,70	0,58
University of Maryland-College Park	448	434	250	26510	6480	632	41,7	15	12	1,00	1,00	1,00	1,00	1,00
University of Massachusetts Amherst	349	133	138	22330	4849	268	28,4	31	49	1,00	0,96	0,96	0,91	1,09
University of Miami	459	233	278	10556	2645	181	28,4	83	56	0,51	0,40	0,80	0,35	0,51
University of Michigan-Ann Arbor	956	817	508	27287	13466	857	60,4	1	2	1,00	1,00	1,00	1,00	1,00
University of Minnesota-Twin Cities	668	739	609	30115	5621	734	49,9	13	5	0,85	0,64	0,76	0,41	0,85
University of Missouri-Columbia	319	161	137	24251	4532	367	20,1	30	48	0,98	0,98	1,00	0,99	0,97
University of Nebraska-Lincoln	206	197	118	17878	3216	246	21,9	63	62	0,88	0,79	0,89	0,96	0,62
University of New Hampshire-Main Campus	155	147	71	13351	1690	58	16,2	84	105	1,00	1,00	1,00	1,00	1,00
University of New Mexico-Main Campus	260	191	107	19167	3796	202	21,8	47	68	0,97	0,97	1,00	0,97	0,97
University of North Carolina at Chapel Hill	723	505	254	18078	6143	495	43,2	26	18	1,00	0,99	0,99	0,40	1,84
University of Notre Dame	332	116	240	9212	2828	210	21,1	86	69	0,98	0,61	0,62	0,47	1,01
University of Oklahoma Norman Campus	275	115	142	18367	3843	218	16	50	80	1,00	0,77	0,77	0,66	1,04
University of Oregon	261	90	121	20118	3323	170	18,7	47	90	0,99	0,87	0,88	0,82	0,99
University of Pennsylvania	1085	704	944	11871	10413	514	58,4	10	10	0,89	0,68	0,76	0,54	0,89
University of Pittsburgh-Pittsburgh Campus	466	690	323	18426	8473	479	42,2	14	20	1,00	0,85	0,85	1,06	0,71
University of Rhode Island	117	99	97	13280	1929	89	16,6	84	102	1,00	0,98	0,98	1,08	0,86
University of Rochester	314	316	157	6461	3473	265	30,6	81	45	0,69	0,69	1,00	0,69	0,69
University of South Carolina-Columbia	295	133	133	22619	4784	279	20,8	31	58	1,00	1,00	1,00	1,00	1,00
University of South Florida-Main Campus	310	278	169	28780	7104	270	21,1	10	57	1,00	1,00	1,00	1,00	1,00
University of Southern California	1436	392	460	18849	12851	634	38,3	3	13	1,00	0,91	0,91	0,84	1,15
University of Utah	342	285	183	21657	6114	339	34,3	20	41	0,87	0,81	0,93	0,90	0,70
University of Vermont	190	101	116	10987	861	62	15,6	91	108	0,83	0,68	0,83	0,46	1,10
University of Virginia-Main Campus	376	345	230	15513	6915	393	36	22	32	1,00	0,96	0,96	0,76	1,17

Universities	Inputs			Outputs				Dominating		Efficiency			Reallocation	
	Expenses ($ million)			Enrollment (FTE)		Research		sets		scores			of budget for	
	Instr	Res	A&I	Undergr	Grad	Doc	Pub	$\|D^i\|$	$\|D^r\|$	TE^d	TE^c	C^oE	Instr	Res
University of Washington-Seattle Campus	1053	890	516	29247	12081	708	59,2	2	5	0,71	0,70	0,99	0,63	0,77
University of Wisconsin-Madison	512	931	252	27872	6270	813	52	13	3	1,00	1,00	1,00	1,00	1,00
University of Wyoming	165	90	83	9300	1462	72	14,6	96	115	0,90	0,85	0,94	0,94	0,69
Utah State University	167	134	87	20080	2015	94	13,7	52	113	1,00	1,00	1,00	1,00	1,00
Vanderbilt University	843	439	191	6814	3880	273	35,1	73	40	0,57	0,57	1,00	0,57	0,57
Virginia Commonwealth University	338	148	137	21703	5080	333	21	28	53	1,00	0,97	0,97	0,94	1,03
Virginia Polytechnic Institute and State U.	306	330	138	24847	5887	469	25,1	20	31	1,00	1,00	1,00	1,00	1,00
Wake Forest University	184	171	619	4639	1625	57	17,4	115	97	0,63	0,58	0,91	0,63	0,52
Washington State University	234	226	151	21399	4487	203	21	36	71	1,00	0,89	0,89	1,24	0,52
Washington University in St Louis	1254	495	269	6934	3565	251	44,9	77	25	1,00	0,94	0,94	0,45	2,17
Wayne State University	316	187	167	15873	4821	229	19,2	46	75	0,90	0,81	0,91	0,92	0,63
Yale University	1288	506	513	6863	6370	390	60	34	7	1,00	0,70	0,70	0,45	1,35
Yeshiva University	221	257	159	2593	1463	129	21,1	122	80	0,55	0,55	1,00	0,55	0,55

Table 2.7: Data on US universities in 2012

Universities	Efficiency			Decentralized model						Centralized model				
				Shadow prices				Weights		Shadow prices			Weights	
	TE^d	TE^c	C^oE	p^I	p^R	P^I	P^R	w^I	w^R	p	P^I	P^R	w^I	w^R
Arizona State University	1,00	1,00	1,00	1,42	3,12	0,78	0,00	0,61	0,39	1,89	0,78	0,00	0,76	0,24
Auburn University	0,84	0,78	0,94	1,89	7,54	1,30	2,52	0,34	0,66	3,99	2,76	0,30	0,71	0,29
Boston College	0,85	0,76	0,90	3,43	8,63	0,00	4,62	0,43	0,57	5,24	0,00	2,80	0,65	0,35
Boston University	0,99	0,95	0,96	0,75	2,71	3,01	0,00	0,73	0,27	1,12	3,01	0,00	0,89	0,11
Brandeis University	1,00	1,00	1,00	6,71	4,23	0,00	16,92	0,46	0,54	6,06	16,70	0,22	0,85	0,15
Brigham Young University-Provo	0,89	0,85	0,95	1,77	6,36	2,50	3,85	0,57	0,43	2,19	4,90	1,46	0,84	0,16
Brown University	0,88	0,76	0,87	1,74	6,98	0,00	3,54	0,23	0,77	3,38	0,00	3,72	0,44	0,56
California Institute of Technology	0,93	0,89	0,96	1,62	3,04	0,00	6,46	0,17	0,83	2,44	0,00	6,46	0,25	0,75
Carnegie Mellon University	0,69	0,62	0,90	1,29	3,73	0,00	3,47	0,22	0,78	2,62	0,00	2,44	0,45	0,55
Case Western Reserve University	0,58	0,53	0,92	3,29	1,25	0,11	4,89	0,46	0,54	2,08	0,07	4,94	0,29	0,71
Clemson University	1,00	1,00	1,00	4,42	2,26	4,53	4,52	0,66	0,34	3,55	3,63	5,41	0,53	0,47
Colorado State University-Fort Collins	0,91	0,87	0,95	3,54	1,73	6,79	0,13	0,80	0,20	2,68	6,72	0,20	0,69	0,31
Columbia University in the City of New York	0,89	0,81	0,91	0,33	1,26	1,31	0,00	0,57	0,43	0,57	1,31	0,00	0,81	0,19
Cornell University	0,93	0,90	0,97	0,74	2,94	0,00	1,03	0,18	0,82	1,03	0,00	2,26	0,25	0,75
Dartmouth College	0,59	0,57	0,96	6,14	3,22	1,53	0,00	0,73	0,27	4,64	1,53	0,00	0,62	0,38
Drexel University	0,82	0,78	0,96	4,58	1,33	1,33	0,00	0,92	0,08	3,66	1,33	0,00	0,78	0,22
Duke University	0,65	0,62	0,95	0,43	0,85	0,00	1,72	0,20	0,80	0,64	0,00	1,72	0,30	0,70
Emory University	0,52	0,51	0,98	2,63	0,80	0,75	0,05	0,82	0,18	1,84	0,69	0,12	0,60	0,40
Florida State University	1,00	1,00	1,00	2,72	1,58	0,00	6,32	0,46	0,54	2,35	4,93	1,39	0,72	0,28
George Mason University	1,00	1,00	1,00	1,95	5,69	0,58	7,24	0,34	0,66	2,74	7,82	0,00	0,89	0,11
George Washington University	0,98	0,97	0,99	3,65	1,21	1,21	0,00	0,91	0,09	1,77	3,69	0,00	0,87	0,13
Georgetown University	0,41	0,39	0,95	2,08	4,45	0,00	1,11	0,44	0,56	2,77	0,00	1,11	0,59	0,41
Georgia Institute of Technology-Main Campus	0,99	0,86	0,87	3,04	0,90	0,00	3,61	0,44	0,56	1,54	0,00	3,61	0,22	0,78
Harvard University	0,98	0,98	1,00	0,30	0,30	0,00	0,98	0,16	0,84	0,30	0,00	0,98	0,16	0,84
Indiana University-Bloomington	1,00	1,00	1,00	1,15	4,59	0,00	4,00	0,31	0,69	1,56	4,59	0,00	0,91	0,09
Indiana University-Purdue University-Ind.	0,89	0,87	0,97	3,65	1,21	1,21	0,00	0,89	0,11	1,46	4,82	0,00	0,86	0,14
Iowa State University	0,98	0,97	0,99	1,93	1,51	0,63	5,42	0,32	0,68	1,75	0,57	5,48	0,29	0,71
Johns Hopkins University	0,78	0,74	0,95	0,68	0,30	1,21	0,00	0,80	0,20	0,51	1,21	0,00	0,67	0,33
Kansas State University	0,88	0,85	0,97	3,97	2,12	8,48	0,00	0,83	0,17	3,18	8,48	0,00	0,74	0,26
Kent State University at Kent	1,00	1,00	1,00	3,15	12,59	0,00	10,27	0,29	0,71	7,94	3,15	0,00	0,91	0,09
Lehigh University	1,00	1,00	1,00	6,24	3,86	9,09	2,36	0,83	0,17	5,64	8,20	3,44	0,75	0,25
Louisiana State University	0,75	0,73	0,98	2,65	1,39	2,15	3,42	0,54	0,46	2,02	0,60	4,97	0,33	0,67
Massachusetts Institute of Technology	0,53	0,52	1,00	0,38	0,59	0,00	1,53	0,13	0,87	0,52	0,00	1,53	0,17	0,83
Michigan State University	0,93	0,91	0,98	1,47	0,79	0,00	3,16	0,47	0,53	1,21	0,00	3,16	0,39	0,61
Montana State University	0,85	0,77	0,90	8,49	3,67	0,00	14,69	0,38	0,62	5,64	0,00	14,69	0,25	0,75
New York University	0,97	0,95	0,98	1,06	0,43	0,00	0,94	0,66	0,34	0,67	0,27	1,45	0,47	0,53
North Carolina State University at Raleigh	0,89	0,87	0,98	2,11	1,34	0,00	4,78	0,42	0,58	1,80	0,00	4,78	0,36	0,64
Northeastern University	0,90	0,89	0,99	2,68	6,35	2,53	0,00	0,70	0,30	3,92	1,59	0,00	0,82	0,18
Northwestern University	0,90	0,86	0,96	0,64	1,06	0,00	2,31	0,21	0,79	0,93	0,00	2,03	0,30	0,70

Universities	Efficiency			Decentralized model						Centralized model				
				Shadow prices				Weights		Shadow prices			Weights	
	TE^d	TE^c	C^oE	p^I	p^R	P^I	P^R	w^I	w^R	p	P^I	P^R	w^I	w^R
Ohio State University-Main Campus	1,00	1,00	1,00	1,59	0,54	0,54	0,00	0,87	0,13	0,77	2,15	0,00	0,81	0,19
Oregon State University	0,80	0,76	0,95	5,22	1,82	3,60	0,14	0,81	0,19	2,60	1,79	5,49	0,40	0,60
Pennsylvania State University-Main Campus	0,65	0,52	0,81	0,38	1,52	0,00	0,54	0,23	0,77	0,53	0,00	1,49	0,32	0,68
Princeton University	0,88	0,88	0,99	1,01	4,03	0,00	1,49	0,20	0,80	1,05	0,00	4,03	0,21	0,79
Purdue University-Main Campus	0,96	0,94	0,98	0,87	3,48	0,00	2,06	0,28	0,72	1,29	0,00	3,48	0,41	0,59
Rensselaer Polytechnic Institute	0,90	0,89	0,99	2,85	5,12	0,00	11,38	0,21	0,79	3,87	0,00	11,38	0,28	0,72
Rice University	0,86	0,80	0,93	2,27	9,08	0,00	6,38	0,29	0,71	3,34	0,00	9,08	0,43	0,57
Rutgers University-New Brunswick	0,92	0,87	0,95	0,67	2,69	0,01	1,67	0,29	0,71	1,05	0,02	2,56	0,45	0,55
Saint Louis University-Main Campus	0,90	0,87	0,97	2,52	10,10	0,00	8,52	0,30	0,70	3,11	0,00	10,10	0,38	0,62
San Diego State University	1,00	1,00	1,00	8,53	3,41	3,41	0,00	0,99	0,01	3,59	0,00	13,64	0,35	0,65
Southern Methodist University	0,95	0,95	1,00	8,10	12,27	0,28	2,79	0,66	0,34	8,62	0,29	2,77	0,70	0,30
Stanford University	0,83	0,75	0,90	0,35	1,30	0,00	0,35	0,22	0,78	0,53	0,00	1,40	0,33	0,67
Stony Brook University	0,76	0,72	0,95	2,27	5,03	1,23	0,37	0,64	0,36	2,83	1,54	0,21	0,80	0,20
SUNY at Albany	0,71	0,64	0,90	6,61	1,71	1,71	0,00	0,80	0,20	3,00	4,69	0,00	0,64	0,36
Syracuse University	0,71	0,66	0,93	3,62	6,70	0,62	1,06	0,64	0,36	4,20	1,01	0,66	0,78	0,22
Temple University	0,76	0,72	0,95	2,58	4,91	0,96	0,26	0,69	0,31	3,07	1,06	0,16	0,80	0,20
Texas A & M University-College Station	1,00	1,00	1,00	1,64	0,73	2,64	0,29	0,77	0,23	1,21	0,00	2,93	0,37	0,63
Texas Tech University	0,98	0,98	0,99	6,93	2,10	1,94	0,16	0,84	0,16	2,65	8,21	0,20	0,80	0,20
The University of Montana	0,98	0,98	1,00	14,76	4,86	0,00	5,08	0,74	0,26	5,75	13,45	6,01	0,69	0,31
The University of Tennessee	0,63	0,62	1,00	1,33	1,09	0,00	3,56	0,38	0,62	1,25	0,00	3,56	0,36	0,64
The University of Texas at Austin	0,94	0,93	0,99	0,58	1,22	2,32	0,00	0,66	0,34	0,85	2,32	0,00	0,76	0,24
The University of Texas at Dallas	1,00	1,00	1,00	4,53	3,02	0,00	12,08	0,36	0,64	3,99	12,08	0,00	0,82	0,18
The University of Texas at San Antonio	0,96	0,94	0,98	9,13	3,04	3,04	0,00	0,92	0,08	4,32	7,70	1,83	0,78	0,22
Tufts University	0,63	0,61	0,97	6,11	1,53	0,00	1,53	0,67	0,33	4,31	0,00	1,53	0,47	0,53
Tulane University of Louisiana	0,70	0,70	1,00	3,39	1,88	0,00	7,52	0,45	0,55	2,82	0,00	7,52	0,37	0,63
University at Buffalo	0,77	0,74	0,96	1,25	4,78	0,63	2,49	0,33	0,67	2,47	1,25	1,29	0,65	0,35
University of Alabama at Birmingham	0,49	0,43	0,88	4,75	1,20	0,00	1,19	0,67	0,33	2,99	0,00	1,19	0,42	0,58
University of Alaska Fairbanks	0,93	0,91	0,97	6,32	3,07	0,00	12,27	0,34	0,66	4,42	0,00	12,27	0,24	0,76
University of Arizona	0,87	0,82	0,94	0,82	2,90	0,00	0,82	0,18	0,82	1,09	0,00	3,26	0,24	0,76
University of Arkansas	0,91	0,88	0,97	4,34	2,35	7,73	0,37	0,83	0,17	3,79	6,75	0,60	0,72	0,28
University of California-Berkeley	0,98	0,97	0,99	0,60	2,14	0,00	0,60	0,20	0,80	0,86	0,00	2,38	0,28	0,72
University of California-Davis	0,88	0,81	0,92	0,59	2,18	0,00	0,59	0,20	0,80	0,84	0,00	2,35	0,28	0,72
University of California-Irvine	0,88	0,84	0,95	0,90	3,59	0,00	1,36	0,23	0,77	1,23	0,00	3,59	0,32	0,68
University of California-Los Angeles	0,63	0,53	0,85	0,33	1,32	0,00	0,55	0,25	0,75	0,47	0,00	1,32	0,35	0,65
University of California-Riverside	0,97	0,96	0,99	2,28	8,94	4,06	0,00	0,44	0,56	3,29	5,86	3,27	0,64	0,36
University of California-San Diego	0,72	0,70	0,97	0,51	0,84	0,00	2,03	0,16	0,84	0,70	0,00	2,03	0,23	0,77
University of California-Santa Barbara	0,95	0,92	0,97	1,78	6,57	0,00	1,78	0,22	0,78	2,68	0,00	7,12	0,32	0,68
University of California-Santa Cruz	1,00	1,00	1,00	2,68	10,07	0,00	2,68	0,20	0,80	3,67	0,00	10,73	0,27	0,73
University of Central Florida	1,00	1,00	1,00	3,09	7,01	1,14	0,61	0,53	0,47	4,24	1,39	0,37	0,72	0,28

Universities	Efficiency			Decentralized model						Centralized model				
				Shadow prices				Weights		Shadow prices			Weights	
	TE^d	TE^c	C^oE	p^I	p^R	P^I	P^R	w^I	w^R	p	P^I	P^R	w^I	w^R
University of Chicago	0,78	0,67	0,86	0,60	2,39	0,00	1,85	0,30	0,70	0,90	0,00	2,39	0,46	0,54
University of Cincinnati-Main Campus	0,86	0,78	0,91	4,52	1,32	1,22	0,10	0,85	0,15	3,22	1,08	0,24	0,63	0,37
University of Colorado Boulder	0,96	0,93	0,97	1,15	2,63	1,63	2,98	0,35	0,65	1,83	2,59	2,02	0,55	0,45
University of Colorado Denver	0,50	0,49	0,98	2,74	1,22	0,00	4,89	0,57	0,43	2,11	0,00	4,89	0,44	0,56
University of Connecticut	0,57	0,49	0,86	1,00	3,99	0,00	2,44	0,24	0,76	2,39	0,00	1,25	0,58	0,42
University of Delaware	0,75	0,68	0,91	1,57	6,26	3,51	0,47	0,53	0,47	2,22	6,10	0,16	0,83	0,17
University of Florida	0,93	0,93	1,00	1,24	0,61	2,43	0,00	0,81	0,19	0,95	0,00	2,43	0,34	0,66
University of Georgia	1,00	1,00	1,00	4,07	1,20	0,00	2,09	0,60	0,40	2,08	0,00	3,61	0,30	0,70
University of Hawaii at Manoa	0,72	0,72	0,99	1,32	3,06	1,92	3,36	0,29	0,71	2,26	2,79	2,49	0,48	0,52
University of Houston	1,00	1,00	1,00	5,60	1,67	1,67	0,00	0,90	0,10	4,10	2,08	0,00	0,76	0,24
University of Illinois at Chicago	0,79	0,77	0,98	1,65	0,86	3,44	0,00	0,87	0,13	1,41	3,44	0,00	0,79	0,21
University of Illinois at Urbana-Champaign	1,00	1,00	1,00	0,69	2,69	0,00	0,69	0,21	0,79	1,60	0,00	0,69	0,48	0,52
University of Iowa	0,83	0,81	0,98	1,01	1,78	0,00	4,05	0,20	0,80	1,37	0,00	4,05	0,27	0,73
University of Kansas	0,73	0,70	0,95	2,98	1,15	1,62	1,15	0,72	0,28	1,70	0,93	3,67	0,41	0,59
University of Kentucky	0,60	0,59	0,98	4,57	1,23	0,00	1,23	0,68	0,32	1,77	0,00	4,35	0,26	0,74
University of Louisville	0,66	0,65	0,98	3,29	1,71	5,15	0,58	0,82	0,18	2,35	3,69	3,16	0,58	0,42
University of Maryland-College Park	0,94	0,91	0,97	0,86	3,10	0,00	0,86	0,20	0,80	1,23	0,00	3,44	0,28	0,72
University of Massachusetts Amherst	0,97	0,95	0,99	1,57	4,01	2,97	3,31	0,49	0,51	2,25	4,25	2,03	0,70	0,30
University of Miami	0,44	0,40	0,91	1,04	4,01	1,63	0,30	0,48	0,52	2,41	0,82	0,18	0,69	0,31
University of Michigan-Ann Arbor	0,96	0,94	0,98	1,45	0,43	0,43	0,00	0,82	0,18	0,61	1,71	0,00	0,74	0,26
University of Minnesota-Twin Cities	0,74	0,64	0,86	0,48	1,80	0,00	0,48	0,17	0,83	1,18	0,00	0,48	0,40	0,60
University of Missouri-Columbia	0,98	0,98	1,00	2,71	1,58	5,13	0,93	0,80	0,20	2,38	4,50	1,41	0,71	0,29
University of Nebraska-Lincoln	0,83	0,78	0,94	5,44	1,87	3,76	0,14	0,80	0,20	2,64	1,83	5,65	0,39	0,61
University of New Hampshire-Main Campus	0,98	0,93	0,95	5,33	2,61	0,00	10,45	0,42	0,58	4,00	0,00	10,45	0,32	0,68
University of New Mexico-Main Campus	0,80	0,79	0,99	3,34	1,74	6,85	0,13	0,82	0,18	2,67	6,78	0,20	0,73	0,27
University of North Carolina at Chapel Hill	0,85	0,77	0,90	0,66	1,59	0,00	2,63	0,24	0,76	1,04	0,00	2,63	0,39	0,61
University of Notre Dame	0,65	0,58	0,89	2,87	5,66	0,00	1,41	0,49	0,51	3,59	0,00	1,41	0,61	0,39
University of Oklahoma Norman Campus	0,85	0,77	0,90	1,83	7,33	3,27	0,97	0,50	0,50	4,33	1,26	0,57	0,70	0,30
University of Oregon	0,87	0,82	0,95	2,07	8,27	5,52	0,00	0,62	0,38	2,71	8,27	0,00	0,88	0,12
University of Pennsylvania	0,70	0,63	0,90	0,56	1,42	0,00	0,36	0,31	0,69	0,90	0,00	0,36	0,50	0,50
University of Pittsburgh-Pittsburgh Campus	0,92	0,85	0,92	2,63	0,66	0,44	0,38	0,70	0,30	1,28	0,69	0,74	0,42	0,58
University of Rhode Island	0,98	0,98	1,00	11,44	3,11	3,11	0,00	0,84	0,16	3,42	12,45	0,00	0,83	0,17
University of Rochester	0,63	0,62	0,98	1,24	2,48	0,00	4,95	0,20	0,80	1,86	0,00	4,95	0,30	0,70
University of South Carolina-Columbia	1,00	1,00	1,00	1,74	6,94	1,89	1,95	0,39	0,61	2,40	1,45	5,49	0,46	0,54
University of South Florida-Main Campus	0,91	0,87	0,96	2,33	1,29	4,39	0,76	0,75	0,25	1,84	4,06	1,09	0,64	0,36
University of Southern California	0,92	0,89	0,96	0,43	1,41	1,15	0,56	0,58	0,42	0,64	1,70	0,00	0,87	0,13
University of Utah	0,84	0,81	0,96	3,09	1,20	2,93	0,08	0,82	0,18	2,40	2,27	0,15	0,63	0,37
University of Vermont	0,72	0,66	0,91	2,40	9,59	0,00	4,55	0,23	0,77	2,88	0,00	9,59	0,28	0,72
University of Virginia-Main Campus	0,95	0,94	0,98	1,02	1,80	0,00	4,10	0,20	0,80	1,39	0,00	4,10	0,27	0,73

Universities	Efficiency			Decentralized model						Centralized model				
				Shadow prices				Weights		Shadow prices			Weights	
	TE^d	TE^c	C^oE	p^I	p^R	P^I	P^R	w^I	w^R	p	P^I	P^R	w^I	w^R
University of Washington-Seattle Campus	0,67	0,66	0,97	0,40	1,49	0,00	0,40	0,21	0,79	0,58	0,00	1,58	0,31	0,69
University of Wisconsin-Madison	0,89	0,88	0,99	0,57	1,15	0,00	2,30	0,15	0,85	0,95	0,00	2,30	0,25	0,75
University of Wyoming	0,83	0,80	0,96	5,58	2,87	7,98	1,22	0,81	0,19	3,88	9,85	1,64	0,75	0,25
Utah State University	0,87	0,82	0,94	4,43	2,51	1,39	8,65	0,44	0,56	3,57	1,12	8,92	0,36	0,64
Vanderbilt University	0,41	0,35	0,86	0,66	2,02	0,00	2,64	0,29	0,71	1,13	0,00	2,64	0,49	0,51
Virginia Commonwealth University	0,98	0,97	0,99	1,56	3,79	2,96	3,29	0,48	0,52	2,24	4,24	2,00	0,69	0,31
Virginia Polytechnic Institute and State U.	1,00	1,00	1,00	2,73	1,26	3,66	1,37	0,69	0,31	1,97	2,88	2,14	0,51	0,49
Wake Forest University	0,45	0,44	0,99	3,99	3,46	1,00	0,00	0,70	0,30	3,74	1,00	0,00	0,67	0,33
Washington State University	0,91	0,88	0,97	5,75	1,59	1,47	0,12	0,81	0,19	2,14	6,21	0,16	0,74	0,26
Washington University in St Louis	0,58	0,49	0,84	0,48	1,66	0,00	1,93	0,31	0,69	0,82	0,00	1,93	0,53	0,47
Wayne State University	0,86	0,81	0,95	4,07	1,45	2,22	0,11	0,85	0,15	3,21	1,75	0,24	0,67	0,33
Yale University	0,78	0,59	0,76	0,42	1,69	0,00	1,07	0,28	0,72	0,60	0,00	1,69	0,40	0,60
Yeshiva University	0,51	0,50	0,97	2,63	1,53	0,91	5,20	0,37	0,63	2,04	0,70	5,41	0,29	0,71

Table 2.8: Shadow prices for models with price restrictions .

Chapter 3

Is there a prison size dilemma? An empirical analysis of output-specific economies of scale

Abstract

We advocate a nonparametric multi-output framework to estimate output-specific economies of scale and we apply this model to male prisons in England and Wales over the sample period 2009-2012. To estimate output-specific returns to scale in prisons, we consider not only the cost-per-place, but also qualitative outputs such as purposeful out-of-cell activity and successful reintegration. Furthermore, we introduce environmental heterogeneity using the characteristics of the prisoners. England and Wales offers a unique example to study economies of scale in prisons as the UK has started to build new super-size prisons in order to replace the most outdated prisons.[1]

[1] This chapter is based on joint work with Marijn Verschelde (IÉSEG School of Management) and Richard Simper (University of Nottingham). We would like to thank Laurens Cherchye, Bram De Rock, Victor Podinovski, the participants of the 14th European Workshop on Efficiency and Productivity Analysis and of the 5th workshop on efficiency and productivity analysis for their insightful comments and suggestions.

3.1 Introduction

Prisoner numbers are on the rise for decades in the US and many European countries.[2] The increasing prisoner population puts the existing United Kingdom (UK) Criminal Justice System (CJS) under stress and forces policy makers to reconsider the limits both from a cost-per-place and from social perspectives (i.e. providing humane incarceration with prospects for reintegration into society when released). Current public policy mainly consists of building new prisons, reshaping existing prisons and putting less convicts behind bars.[3]

This chapter focuses on what we call the potential '*prison size dilemma*'. Public policy makers could consider returns to scale from either a cost-per-place or a social viewpoint, but these viewpoints lead to conflicting opinions on the optimal scale size. We empirically test whether the optimal scale size of a prison differs when the focus is either on costs-per-place, quality of life in prison or successful reintegration.[4] In particular, we study economies of scale of a sample of male prisons in England and Wales, by using publicly available data collected by the Ministry of Justice (MoJ).

Our empirical analysis is timely and warranted as the building of the first titan prison in the UK has started.[5] The Labour government in 2007 was forced to abandon 3 titan prisons which would provide up to 2,500 places in five units of approximately 500 offenders[6].

[2]See Levitt (1996) and Campbell et al. (2015) for a discussion on US mass incarceration and e.g. the National Audit Office (2013, p. 14) for prisoner figures for England and Wales.

[3]The latter has been considered both from an operational and a deterrence point of view. See for example the National Audit Office (2013, p. 14). for the operational point of view. For the deterrence point of view, a well-established literature, inspired by the seminar work of Becker (1968), shows positive, but highly accelerating diminishing returns from more incarcerations to reduce crime (See e.g. Levitt (1996), Buonanno and Raphael (2013), Di Tella and Schargrodsky (2013), Vollaard (2013), Hansen (2015) and references therein).

[4]While there is evidence for a deterrence effect of harsh prison conditions (e.g. Katz et al. (2003)), we consider efforts to foster reintegration and quality of prison life as 'goods'. A large criminological literature studies violence in prison and shows both effects from importation of violence (e.g. Mears et al. (2013)) and deprivation due to among others poor prison management (e.g. Sykes (1958) and McCorkle et al. (1995)). In England and Wales, it is now fully acknowledged that the high proportion of offenders that re-offend after discharge is costly to the tax payers and society (see e.g. Ministry of Justice (2011)and Ministry of Justice (2013)).

[5]The name titan refers to not just a single large prison but one consisting of hubs. "*Hub prisons would be large establishments of between 2,500 - 3,000 places. They would be designed to be operated as a number of semi-autonomous units sharing a common site and set of services; provide operational flexibility to respond to changes in the size and profile of the prison population*" Lockyer (2013, page 6)

[6]This was due to political pressure and against the main recommendations to go ahead by the Carter Review (2007). However, Lord Carter did note that "*there are some operational challenges associated with large prisons, including the possibility of large scale disturbance, the difficulty in meeting the needs of specific groups of prisoners*

However, the rejection of building these titan prisons was reversed in 2011 under the next UK government - a Conservative/Liberal coalition - where the building of the first titan prison based in Wrexham, Wales was agreed to begin.[7] Renewing and rescaling the prison estate is part of the strategy of the National Offender Management Service (NOMS), which covers both the prison and probation systems in England and Wales, to reduce costs. The modernization of the prison estate includes the closure of old and inefficient prisons, which will be replaced by new large prisons and housing blocks.[8]

The cost reduction strategy of the NOMS, initiated in 2010, also involved a reduction of input waste within the system. Furthermore, the NOMS aimed to introduce more competition by privatization and re-tendering of prisons that were already tendered to the private sector. Rogge et al. (2015) document that there is little empirical support for large cost savings contracting-out prison service to private-run organizations. In our study, we analyze the optimal scale size of prisons.

From a methodological perspective, we advocate a framework that is specially tailored to analyze the multidimensional prison production process. In particular, we propose a DEA-based methodology that fully acknowledges that returns to scale can differ between the different (qualitative) dimensions of production.

We build on the work of Cherchye et al. (2013), who introduce a multi-output methodology that recognizes that each output is characterized by its own production technology. Starting from this multi-output methodology, we will be able to estimate output-specific returns to scale.

An attractive feature of the methodology is that it is nonparametric: there is no need to assume a specific functional representation of the production technology. This is warranted for public sector applications as public firms operate in non-competitive markets and can

(e.g. female and young offenders) and the management complexities associated with a large staff complement and challenges of managing a number of potentially different prisoner segments on the same site" page 38.

[7]The first titan prison would cost £212 million and would be operated by Her Majesty's Prison Service (HMPS) where it is to outsource up to 34% of services to private and voluntary sector.

[8]On January 10, 2013, The Ministry of Justice announced the closure of four prisons and partial closure of three prisons. In total, 2,614 places were closed. An announcement on September 4, 2013 showed an even more drastic change of the prison landscape as in the period 2010-2014, the prison (planned) closures consist in total of 6,382 places and total gained places in micro-prisons (housing blocks) or new large prisons are up to 5,945. For more information, see URL: https://www.gov.uk/government/news/modernisation-of-the-prison-estate.

have a complex structure of public production. Consequently, the imposition of a parametric functional relationship can be intricate. Instead, a minimum set of production axioms is used to test for output-specific economies of scale.

In the context of prisons, we argue that it is crucial to consider output-specific returns to scale. We take three output objectives into account. Naturally, we consider the incarceration of convicts as one of the main outputs of a prison. Besides incarcerating convicts, we consider in our study also qualitative outputs including the provision of a humane prison environment and successful reintegration. In the empirical analysis, we select proxies that in our opinion best reflect these output objectives.

A common motivation for large prisons is a reduction of the cost-per-place. Meanwhile, opponents fear little prospects for reintegration and low quality of life in large-scale prisons.[9] In fact, the HM Chief Inspector of Prisons (2009) and The National Audit Office (2013) provide support in England and Wales for a better performance in smaller prisons. Surveys show that prisoners tend to be more engaged in smaller establishments.[10] Moreover small prisons do on average better in independent inspections and in the NOMS's performance ratings, which take reintegration and quality of life in prison into account. By contrast, Lockyer (2013) argues that the age and not the size of a prison determines the performance of a prison. In our opinion there is a need for further research on the relation between prison size and the multiple facets of performance.

The above-mentioned studies do not control for prison(er) characteristics. However, Ruggiero (2000) emphasizes that environmental variables have a considerable impact on the provision of public services and that without controlling for these environmental factors the estimates of returns to scale will be biased. We advocate a methodology that explicitly takes into account environmental heterogeneity. For example, we control in our study for the inflow of prisoners in particular establishments. Specifically, we include the predicted rate of re-offending in an establishment. The rate of re-offending is estimated at prison level by

[9] For example, Liebling (2004) questions the moral performance of the so called '*Titan*' prisons that could hold over 2,500 prisoners.

[10] In particular, prisoners in small prisons (lower than 400 prisoners) in comparison to large prisons (over 800 prisoners) have significantly higher agreement that the prison addresses well drugs problems, prisoner safety, quality of policing and security, levels of organization and consistency, staff professionalism, quality for the vulnerable and relationships with staff.

the Ministry of Justice, based on prisoner-level data on social background, ethnicity, crime type, etc. We assume that prisons with a higher predicted rate of re-offending operate in a harsher production environment.

Furthermore, the proposed methodology distinguishes between discretionary and non-discretionary output variables. We therefore measure the performance of prisons only with respect to the output variables that the prison management controls and actually wants to maximize. Examples of non-discretionary variables in our application are the size of the average prison population and the yearly number of discharges.

To our knowledge, we posit an original estimation strategy that adequately models the multidimensional prison production process. The advocated methodology is tailored to all specificities of the prison production process and enables us to meaningfully answer the prison size dilemma, by using publicly available data. Moreover, we discuss in detail how public policy makers can further refine the analysis by adding information on the allocation of expenses to particular outputs.

The remainder of the chapter is structured as follows. Section 2 explains the nonparametric multi-output methodology. Section 3 discusses the data and the empirical model and Section 4 discusses the results. Section 5 concludes.

3.2 Methodology

3.2.1 Basic concepts

To set the stage, we first intuitively introduce our ideas and discuss some relevant literature. Next, we formalize the methodology in Section 3.2.2 to 3.2.5.

Returns to scale The concept of returns to scale is directly related to the most productive scale size. A Decision Making Unit (DMU) that is situated on the constant returns to scale technology, is considered to operate on its most productive scale size (Banker, 1984). A DMU which is not situated on its most productive scale size, can improve its productivity by resizing the scale of its operations. The type of returns to scale can be interpreted as the direction of change necessary to achieve its most productive scale size.[11] In particular, in a constant returns to scale technology, a proportional increase in input $\mathbf{X}$ gives a proportional increase in output $\mathbf{y}$. Consequently, a property of the constant returns to scale technology is that the average productivity remains constant, for a given input and output mix. Increasing returns to scale indicate that the most productive scale size of a DMU is situated at a larger size. Similarly, decreasing returns to scale indicate that the DMU should decrease the scale of its operations to achieve the optimal scale size. The type of returns to scale is therefore very useful information for the operational manager, indicating how rescaling the operation can improve average productivity and reduce the average cost.[12]

Output-specific production technology To estimate output-specific returns to scale, we build on the work of Cherchye et al. (2013), who introduce a multi-output methodology that recognizes that each output is characterized by its own production technology. The output-specific production technologies remain linked through the use of joint inputs. In Sections 3.2.3 and 3.2.4 we will focus on the production process of one particular output, to come back to the multi-output production process in Section 3.2.5. We extend Cherchye et al. (2013) by including alternative returns to scale assumptions in the methodology. Furthermore, we include output-specific environmental variables in the methodology. Since we are able to work with output-specific production technologies, we can estimate output-specific returns-to-scale, controlling for output-specific environments.

At this point, it might be worth to note that our approach bears some analogy to Cook and Zhu (2011), who also allow returns to scale type behavior to be different for one output subgroup than for another, by using the notion of component technologies. However, we

[11] We estimate global returns to scale. Podinovski (2004a) and Podinovski (2004b) make the distinction between local and global returns to scale. In a convex production technology, these concepts are identical. In non-convex technologies there is a difference. The type of local returns to scale is indicative of the type of resizing that should lead to immediate improvements of the average productivity. The type of global returns to scale is indicative of the direction of change necessary to achieve maximum average productivity. Since we work in a setting with relaxed convexity assumptions, we are only able to estimate global returns to scale.

[12] We estimate qualitative characterizations of returns to scale, such as increasing, decreasing or constant returns to scale. There is a different strand of DEA literature which is directed to quantitative directions of returns to scale. For example Podinovski and Forsund (2010) and Atici and Podinovski (2012) analyze a class of mixed partial elasticity measures. These measures indicate the elasticity of response of a subset of outputs with respect to marginal changes of a subset of inputs. This approach applies to polyhedral technologies. However, since we work in a setting which is very non-smooth because of the relaxed convexity assumptions, we do not pursue the study of marginal changes.

offer an axiomatic approach to the estimation of output-specific returns to scale.

Returns to scale estimation To estimate returns to scale, we follow a method discussed by Podinovski (2004a) and Podinovski (2004b), based on Kerstens and Vanden Eeckaut (1999).[13] This method is based on the principle of goodness-of-fit. The goodness-of-fit with respect to a particular production technology is measured as the distance of an observed input-output combination to the boundary of the technology. To determine the most appropriate returns to scale assumption, we asses the goodness-of-fit of several production technologies, each based on an alternative returns to scale assumption.

Scale efficiency Finally, to estimate scale efficiency, we follow Banker (1984) in comparing the distance of an observation to the constant returns to scale technology with the distance to the variable returns to scale technology. The constant returns to scale technology consists of DMUs that operate on their most productive scale size. In contrast to the constant returns to scale technology, the variable returns to scale technology makes no assumption at all on the prevailing returns to scale. The distance between the two respective technologies indicates scale efficiency.

Robust methodology Since the estimation of returns to scale is sensitive to outliers, we combine our methodology with the robust order-m method, as introduced by Cazals et al. (2002), discussed in Daraio and Simar (2007a) and elaborated for convex technologies in Daraio and Simar (2007b). In particular, we repeatedly draw a sample of size m (with replacement) among the DMUs in the environment of DMU n. For each random draw, we estimate the efficiency and the returns to scale of DMU n on the basis of the sample of potential comparison partners for DMU n. Finally, the robust efficiency measure is computed as the average over all draws. This procedure allows us to report the statistical significance of the estimations, which is based of the percentage of draws that leads to a particular returns to scale estimate. However, for ease of notation, we describe the methodology without order-m robustification.

[13] Both for observations on the production frontier and below the frontier, it is possible to determine the returns to scale. In the second case, we actually estimate the returns to scale of the projection of the (inefficient) observation on the production frontier.

3.2.2 Notational preliminaries

We observe data for N DMUs. Suppose that each DMU n ($1 \leq n \leq N$) uses input $\mathbf{x}_n = (x_n^1, \ldots, x_n^L)$ to produce output $\mathbf{y}_n = (\mathbf{y}_n^1, \ldots, \mathbf{y}_n^R)$ and is situated in environment $\mathbf{z}_n = (z_n^1, \ldots, z_n^K)$. Note that output $\mathbf{y}_n^r$ can be a set of outputs having a common production technology.

Following Cherchye et al. (2013) and Cherchye et al. (2015b), we consider a separate production technology for each output. Importantly, we account for interdependencies between the different technologies through jointly used inputs.
In particular, we distinguish between output-specific, joint and subjoint inputs. Output-specific inputs can be allocated to the production of particular outputs. We use α_l^r, with $\sum_{r=1}^{R} \alpha_l^r$=1, to represent the fraction of input l that is used to produce output r. Next, joint (or public) inputs simultaneously benefit the production of all outputs. Subjoint inputs also figure as joint inputs, but only for a subset of outputs. The use of joint and subjoint inputs therefore makes the output-specific production processes interdependent. Note that the methodology can also be applied when all inputs are joint (as is the case in our empirical analysis).

We summarize the information on how inputs are allocated to outputs by means of a vector $\mathbf{A}^r$ for each output r. Specifically, $(\mathbf{A}^r)_l = \alpha_l^r$ if input l is output-specific and used to produce output r. Next, $(\mathbf{A}^r)_l = 1$ if input l is joint or sub-joint and used to produce output r. Finally, $(\mathbf{A}^r)_l = 0$ otherwise. The element-by-element product $\mathbf{X}^r$ = $\mathbf{A}^r \odot \mathbf{x}$ captures the input quantities used in the production process of output r.

Next, some environmental variables can influence only a part of the outputs, not all. The vector $\mathbf{B}^r$ captures the environmental variables that are relevant for output r. In particular, $(\mathbf{B}^r)_k = 1$ if environmental variable k is relevant for output r and $(\mathbf{B}^r)_k = 0$ otherwise. Summarizing, the element-by-element product $\mathbf{Z}^r = \mathbf{B}^r \odot \mathbf{z}$ captures the environmental variables that are controlled for in the specification of the technology of output r.

Taken together, the empirical analysis starts from the following data set:

$$S = \{(\mathbf{y}_n^1, \ldots, \mathbf{y}_n^R, \mathbf{X}_n^1, \ldots, \mathbf{X}_n^R, \mathbf{Z}_n^1, \ldots, \mathbf{Z}_n^R) \mid n = 1, \ldots, N\}. \tag{3.1}$$

3.2.3 Output-specific production technology

In this section, we focus on the production technology of output r. For output r, we observe for each DMU n the inputs $\mathbf{X}_n^r$ that are used to produce output $\mathbf{y}_n^r$, in environment $\mathbf{Z}_n^r$. We adopt an output-oriented approach[14] and characterize the production technology of output r by output sets $P^r(\mathbf{X}^r, \mathbf{Z}^r)$, which contains the amount of output $\mathbf{y}^r$ that can be produced with input $\mathbf{X}^r$, in environment $\mathbf{Z}^r$:

$$P^r(\mathbf{X}^r, \mathbf{Z}^r) = \{\mathbf{y}^r | \mathbf{X}^r \text{ can produce } \mathbf{y}^r, \text{ given } \mathbf{Z}^r\}. \tag{3.2}$$

In practice, the true output sets $P^r(\mathbf{X}^r, \mathbf{Z}^r)$ are not observed. A solution is to construct empirical approximations of these output sets on the basis of some standard production axioms.

Axiom 3.1 (Monotone output sets)**.**

$$\mathbf{y}^r \in P^r(\mathbf{X}^r, \mathbf{Z}^r) \text{ and } \mathbf{y}^{r*} \leq \mathbf{y}^r \Rightarrow \mathbf{y}^{r*} \in P^r(\mathbf{X}^r, \mathbf{Z}^r)$$

Axiom 3.2 (Nested output sets)**.**

$$\mathbf{X}^r \leq \mathbf{X}^{r*} \Rightarrow P^r(\mathbf{X}^r, \mathbf{Z}^r) \subset P^r(\mathbf{X}^{r*}, \mathbf{Z}^r)$$

Axiom 3.3 (Convex output sets)**.**

$$\mathbf{y}^r \in P^r(\mathbf{X}^r, \mathbf{Z}^r) \text{ and } \mathbf{y}^{r*} \in P^r(\mathbf{X}^r, \mathbf{Z}^r)$$

$$\Rightarrow \forall \lambda \in [0,1] : \lambda \mathbf{y}^r + (1-\lambda)\mathbf{y}^{r*} \in P^r(\mathbf{X}^r, \mathbf{Z}^r)$$

[14] In this respect we deviate from Cherchye et al. (2013), who follow an input oriented approach and characterize the production technology by input requirement sets.

Axiom 3.4 (Observability means feasibility)**.**

$$(\mathbf{y}^r, \mathbf{X}^r, \mathbf{Z}^r) \in S \Rightarrow \mathbf{y}^r \in P^r(\mathbf{X}^r, \mathbf{Z}^r)$$

Axiom 3.5 (Environmental effect)**.**

$$\mathbf{Z}^{r*} \leq \mathbf{Z}^r \Rightarrow P^r(\mathbf{X}^r, \mathbf{Z}^r) \subset P^r(\mathbf{X}^r, \mathbf{Z}^{r*})$$

Essentially, the first two axioms say that inputs and outputs are freely disposable. Axiom 3.1 states that, if input $\mathbf{X}^r$ can produce output $\mathbf{y}^r$, then it can also produce any lower output $\mathbf{y}^{r*}$. Likewise Axiom 3.2 indicates that more input does not reduce the output. Further, Axiom 3.3 states that, if input $\mathbf{X}^r$ can produce both output $\mathbf{y}^r$ and $\mathbf{y}^{r*}$, then it can also produce any convex combination of these outputs.[15] Axiom 3.4 states that the observed input-output combinations are certainly feasible. Following Ruggiero (1996) we control for environmental variables that affect production. Here we assume that the larger $\mathbf{Z}^r$, the less favorable the production environment. Consequently, Axiom 3.5 implies that a DMU in a more favorable environment should be able to produce at least as much output as any DMU in a less favorable environment.[16]

We add one final axiom, which includes returns to scale in the methodology. We assume either variable returns to scale (vrs), non-increasing returns to scale (nirs), non-decreasing returns to scale (ndrs) or constant returns to scale (crs). We include rts^r in the notation of the output set, indicating which returns to scale assumption we make for output r.

Axiom 3.6 (Output-specific returns to scale)**.**

$$\mathbf{y}^r \in P^r(\mathbf{X}^r, \mathbf{Z}^r, rts^r) \Rightarrow k\mathbf{y}^r \in P^r(k\mathbf{X}^r, \mathbf{Z}^r, rts^r) \text{ for } k \in K(rts^r)$$

[15] We work in a setting with relaxed convexity assumptions: we assume convex output sets, but we do not impose convexity in the input-output space. A growing strand of literature assumes a weaker form of convexity, see Podinovski and Kuosmanen (2011) for an overview. The motivation is that a fully convex production set is not well suited for modeling economies of scale (Petersen (1990) and Bogetoft (1996)).

[16] In the absence of information on the direction of influence of environmental factors, a kernel weighting approach can be used as discussed in detail in Daraio and Simar (2005, 2007). An alternative approach to account for environmental variables in a nonparametric efficiency evaluation, is to conduct a two-step procedure. The first step then computes the nonparametric efficiency estimates, and the second step regresses these nonparametric estimates on the environmental variables. However, this approach involves implicit assumptions that are often problematic. See Simar and Wilson (2007) for an insightful discussion.

where rts^r='vrs', 'nirs', 'ndrs', or 'crs' and

where $K(vrs) = \{1\}$, $K(nirs) = [0, 1]$, $K(ndrs) = [1, \infty)$ and $K(crs) = \mathbb{R}$.

The returns to scale assumption describes the change in output resulting from a proportional change in inputs, in a particular environment. If input $\mathbf{X}^r$ can produce output $\mathbf{y}^r$ in environment $\mathbf{Z}^r$, then $k\mathbf{X}^r$ can produce $k\mathbf{y}^r$ for $k \in K(rts^r)$. Depending on which returns to scale assumption that is made, the potential to scale up or down differs. Variable returns to scale is the weakest assumption, under which the input-output combinations can not be scaled. Under the assumption of non-increasing returns to scale, we can scale down the observations. Similarly, the assumption of non-decreasing returns to scale enables us to scale up the observations. Constant returns to scale is the strongest assumption and allows to scale both up and down.

We define the empirical approximation $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ of the output set as the smallest set that is consistent with Axioms 1-5. This is an application of the minimum extrapolation principle which is commonly used in DEA, see Banker et al. (1984).

Illustrative example Before giving a formal definition of the empirically constructed output set, we illustrate the construction with a single input, single output example, which is depicted in Figure 3.1.

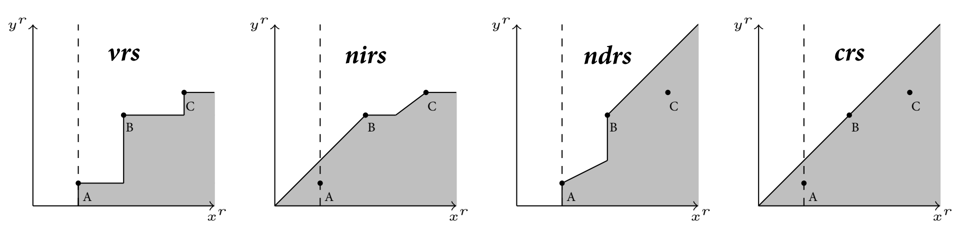


Figure 3.1: Production technology under variable, non-increasing, non-decreasing and constant returns to scale.

We observe the input-output combinations of DMU A, B and C. Assume for simplicity that each DMU is situated in the same environment. The grey area displays the technology set. The relation between the technology set $T^r(rts^r)$ and the output sets $P^r(\mathbf{X}^r, \mathbf{Z}^r, rts^r)$

is the following: $T^r(rts^r) = \{(\mathbf{X}^r, \mathbf{Z}^r, \mathbf{y}^r, rts^r) | \mathbf{y}^r \in P^r(\mathbf{X}^r, \mathbf{Z}^r, rts^r)\}$. Along a vertical line we can therefore read an output set, for a particular input level.

In a first step in the construction, we apply Axiom 3.4, observability means feasibility. This axiom indicates that the technology set is constructed on the basis of the observed input-output combinations A, B and C. In a second step, Axioms 3.1 and 3.2 imply that the input-output combinations to the bottom right of A, B and C are feasible. Since we only have one output in our example, Axiom 3.3 adds no additional information here.[17] In a final step, we include the returns to scale assumption. Figure 3.1 shows the technology sets under the four alternative returns to scale assumptions.

This example consists of three DMUs in the same environment. If this is not the case, only the DMUs which operate in a less favorable environment than DMU n should be used to construct the output set $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$. The construction of the output set therefore starts from the DMUs s with $\mathbf{Z}_n^r \leq \mathbf{Z}_s^r$.

Formal construction Petersen (1990) and Bogetoft (1996) define a scaling function β_s^r, which brings the production of DMU s on a similar scale as the production of DMU n:

$$\beta_s^r(\cdot, rts^r) : \mathbb{R}_0 \to \mathbb{R}_0 \cup \{-\infty\}$$

where

$$\beta_s^r(\mathbf{X}^r, rts^r) = \sup\{\beta \mid \beta \mathbf{X}_s^r \leq \mathbf{X}^r, \beta \in K(rts^r)\}.$$

The scaling parameter $\beta_s^r(\mathbf{X}_n^r, rts^r)$ relates the amount of input of DMU n to the input of DMU s and implies that $\beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{y}_s^r \in P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ for $\mathbf{Z}_n^r \leq \mathbf{Z}_s^r$. The scaling parameter therefore determines to what extent we should scale the output produced by DMU s for the scaled output of DMU s to figure as a benchmark for DMU n. We let $\sup(\emptyset) = -\infty$. In the case that $\beta_s^r(\mathbf{X}_n^r, rts^r)$ equals $-\infty$, it is not possible to compare the output of DMU s with the output of DMU n. We now define the set $C_n^r(rts^r)$ which captures the comparison

[17] In a setting with one output, our technology corresponds to a free disposal hull technology. The free disposal hull (FDH) model (see Deprins et al. (1984) and Tulkens (1993)) does not require convexity, in contrast to the popular Data Envelopment Analysis (DEA) models. We do assume convexity in output space, but not in input-output space.

partners for DMU n with respect to output r as follows:

$$C_n^r(rts^r) = \{s|\mathbf{Z}_n^r \leq \mathbf{Z}_s^r \text{ and } \beta_s^r(\mathbf{X}_n^r, rts^r) > 0\}.$$

In particular, this set consist of all DMUs s which are situated in a less favorable environment as DMU n and which can be rescaled to be compared with DMU n. The empirical output sets $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ are then constructed on the basis of the scaled observations:

$$\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r) = \left\{ \mathbf{y} \middle| \begin{array}{c} \mathbf{y} \leq \sum_{s \in C_n^r(rts^r)} \lambda_s^r \beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{y}_s^r \\ \sum_{s \in C_n^r(rts^r)} \lambda_s^r = 1, \lambda_s^r \geq 0 \end{array} \right\}.$$

This empirical construction satisfies the minimum extrapolation principle, under Axioms 3.1 to 3.6.

Proposition 3.1. $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ satisfies Axioms 3.1 - 3.6. Moreover, for any $P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ that satisfies Axioms 3.1 - 3.6, we have that $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r) \subseteq P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$.

The set $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ thus gives an inner bound approximation of the true output set $P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$, under the given technology axioms.

Now that we have constructed empirical approximations of the output sets, we can define output-specific efficiency. The output-specific efficiency of DMU n can be interpreted as the fit of a particular technology for observation n.

Output-specific technical efficiency Following Banker and Morey (1986), we allow for both discretionary and non-discretionary outputs.[18] We therefore divide the vector $\mathbf{y}_n^r$ in a discretionary and a non discretionary part: $\mathbf{y}_n^r = (\mathbf{y}_{D\,n}^{\ r}, \mathbf{y}_{F\,n}^{\ r})$.

We define the following Farrell (1957) - Debreu (1951) efficiency measure for the production of output r:

[18] We assume that the non-discretionary outputs can be scaled down or up. Although the original Banker and Morey model does not allow scaling, a commonly used version of the model does allow scaling. Syrjänen (2004) extensively discusses that some models require non-discretionary factors to be scale independent, i.e. indices, and some require them to be scale dependent, i.e. volume measures. Syrjänen (2004) proposes a generalized model in which both scale-dependent and scale-independent non-discretionary factors are included. However, we assume in our model that the non-discretionary factors are scale dependent.

$$\hat{\varphi}_n^r(rts^r) = \max\{\varphi | (\varphi \mathbf{y}_{Dn}^{\ r}, \mathbf{y}_{Fn}^{\ r}) \in \hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)\}. \tag{3.3}$$

The measure $\hat{\varphi}_n^r(rts^r)$ captures the distance of DMU n to the boundary of the empirically constructed output set. Stated differently, $\hat{\varphi}_n^r(rts^r)$ indicates the equiproportionate expansion of discretionary output that is certainly feasible, under Axioms 3.1 - 3.6. In general, $1 \leq \hat{\varphi}_n^r(rts^r)$ with $\hat{\varphi}_n^r(rts^r) = 1$ indicating full output-specific technical efficiency. Since $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r) \subseteq P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$, the measure $\hat{\varphi}_n^r(rts^r)$ defines a lower bound for the true, but unobserved measure φ_n^r (with respect to the true, but unobserved output set $P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$).

The measure $\hat{\varphi}_n^r(rts^r)$ is straightforward to compute by a two-step procedure. In a first step, compute the values of the functions $\beta_s^r(\mathbf{X}_n^r, rts^r)$ and the DMUs in the set $C_n^r(rts^r)$. In a second step, the measure $\hat{\varphi}_n^r(rts^r)$ can be computed by solving the following linear programming problem (LP-1):

$$\begin{aligned}
\hat{\varphi}_n^r(rts^r) \quad &= \quad \max_{\varphi_n \geq 0, \lambda_s^r \geq 0} \varphi_n \\
s.t. \quad & \\
\text{(D-1)} \quad & \sum_{s \in C_n^r(rts^r)} \lambda_s^r \beta_s^r(\mathbf{X}_n^r, rts^r) \mathbf{y}_{Ds}^{\ r} \geq \varphi_n \mathbf{y}_{Dn}^{\ r} \\
\text{(D-2)} \quad & \sum_{s \in C_n^r(rts^r)} \lambda_s^r \beta_s^r(\mathbf{X}_n^r, rts^r) \mathbf{y}_{Fs}^{\ r} \geq \mathbf{y}_{Fn}^{\ r} \\
\text{(D-3)} \quad & \sum_{s \in C_n^r(rts^r)} \lambda_s^r = 1.
\end{aligned}$$

Note that linear programming problem (LP-1) also has an intuitive interpretation in terms of revenue efficiency. We refer to Appendix 3.C for the dual formulation.

3.2.4 Output-specific returns to scale

Since each output has its own production technology, we can estimate returns to scale for every individual output. Traditionally, returns to scale are said to be either constant, increasing or decreasing. In a setting of relaxed convexity assumptions, Podinovski (2004a) and Podinovski (2004b) introduce a fourth option, namely sub-constant returns to scale. Sub-constant returns to scale indicate that the most productive scale size can be achieved by either reducing or increasing its scale. The identification of the returns to scale is based on the definition of the variable, non-increasing and non-decreasing returns to scale technologies:

- Constant returns to scale $\Leftrightarrow \hat{\varphi}_n^r(vrs) = \hat{\varphi}_n^r(nirs) = \hat{\varphi}_n^r(ndrs)$
- Decreasing returns to scale $\Leftrightarrow \hat{\varphi}_n^r(vrs) \leq \hat{\varphi}_n^r(nirs) < \hat{\varphi}_n^r(ndrs)$
- Increasing returns to scale $\Leftrightarrow \hat{\varphi}_n^r(vrs) \leq \hat{\varphi}_n^r(ndrs) < \hat{\varphi}_n^r(nirs)$
- Sub-constant returns to scale $\Leftrightarrow \hat{\varphi}_n^r(vrs) < \hat{\varphi}_n^r(ndrs) = \hat{\varphi}_n^r(nirs)$.

For example in Figure 3.2, we estimate that DMU A and D exhibit increasing returns to scale, DMU B constant returns to scale and DMU C decreasing returns to scale.[19]

Output-specific scale efficiency A DMU operating under constant returns to scale is considered to operate on its optimal scale size. The output-specific technical efficiency measure under the assumption of constant returns to scale thus takes into account deviations from the optimal scale size. By contrast, the output-specific technical efficiency measure under variable returns to scale gives each DMU the benefit of the doubt with respect to its scale size. Comparing these efficiency measures therefore gives an indication of the extent to which a DMU deviates from the point of optimal scale of operation. We define a measure of scale efficiency as the ratio of the output-specific technical efficiency measure under constant re-

[19]We could incorporate the information about the estimations of the returns to scale in the efficiency analysis. The idea is to use the intersection of the non-decreasing and the non-increasing returns to scale technology as benchmark technology. Doing so, we give each DMU the benefit of the doubt with respect to its scale size. However, this benchmark technology has more power to identify inefficient behavior than the variable returns to scale technology. Kerstens and Vanden Eeckaut (1998) use the intersection of the increasing returns to scale and decreasing returns to scale technology in a free disposal hull (fdh) technology. The fdh technology imposes strong disposability of inputs and outputs, but without any convexity assumption. We work in a setting with relaxed convexity assumptions: we only assume convex output sets. When convexity in input-output space is imposed, $P(vrs) = P(ndrs) \cap P(nirs)$. Without this assumption, $P(vrs) \subset P(ndrs) \cap P(nirs)$. Therefore, the intersection of these technologies is particularly interesting in a setting with relaxed convexity assumptions.

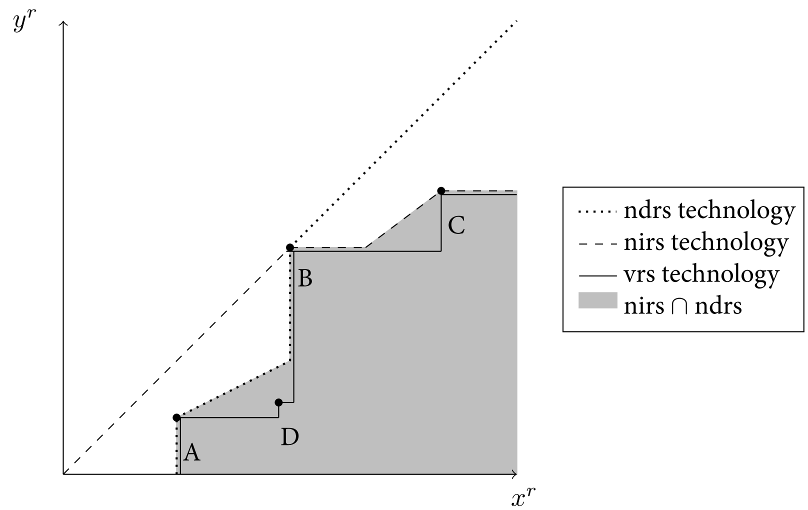


Figure 3.2: Estimation of returns to scale

turns to scale and the measure under variable returns to scale:

$$SE_n^r = \hat{\varphi}_n^r(c)/\hat{\varphi}_n^r(v).$$

3.2.5 Multi-output technical efficiency

Until now we focused on the production process of the individual outputs. However, the production of the individual outputs is linked through the use of joint (and subjoint) inputs. Following Cherchye et al. (2013) we define in this section multi-output efficiency measures that consider the production of all the outputs. An interesting feature of this methodology is that it allows for returns to scale that are specific to individual outputs. The vector **rts** $= (rts^1, \ldots, rts^R)$ captures the returns to scale assumptions rts^r for every output r. We define

$$\hat{\varphi}_n(\mathbf{rts}) = \max\{\varphi | \forall r : (\varphi \mathbf{y}_{D}{}_n^r, \mathbf{y}_{F}{}_n^r) \in \hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)\}. \tag{3.4}$$

In practice, this multi-output technical efficiency measure is computed as follows:

$$\hat{\varphi}_n(\mathbf{rts}) = \min\{\hat{\varphi}_n^1(rts^1), \ldots, \hat{\varphi}_n^R(rts^R)\}. \quad (3.5)$$

3.3 Empirical prison production model

For the empirical analysis of the prison size dilemma, we collected publicly available data on 34 local male prisons in England and Wales, over the book years 2009/10, 2010/11 and 2011/12.[20] We pool the data over the years and obtain 102 observations. By pooling the data, we impose that all observations operate under the same technology, but still allow they can vary with respect to returns to scale, scale efficiency and technical efficiency.

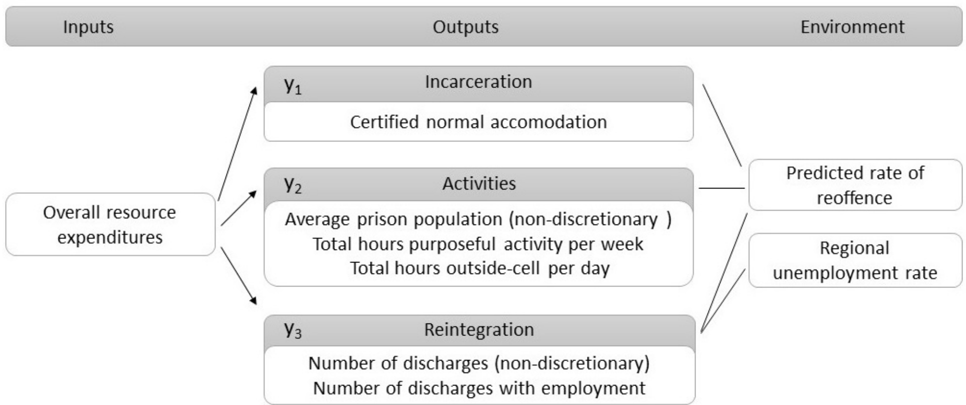
Figure 3.3 shows the empirical prison model we advocate to approximate the true conduct in local prisons in England and Wales, which is by nature multi-dimensional and complex. We presume the local prisons use resources — that come from different sources — to maximize three outputs: (1) Keeping convicts outside society, which we label as "*incarceration*", (2) Providing a humane prison environment and prepare prisoners for reintegration mainly by organizing purposeful and outside-cell activities, (3) Successfully reintegrating discharged prisoners into society. As prisons do not operate in vacuum, we control for both regional and prisoner characteristics, which we allow to be output-specific. As not all aspects of production are controllable for prison management, we distinguish between '*discretionary*' and '*non-discretionary*' output variables. Non-discretionary output variables are non-discretionary for the prison management, but are discretionary for higher-level decision makers. This implies the non-discretionary outputs, in contrast to environmental variables, can be re-scaled to improve scale efficiency.[21]

Table 3.1 shows descriptive statistics for each of the included variables. To approximate the inputs, outputs and environmental heterogeneity, we use proxies that are used in policy making and in our opinion best reflect the true production process:

[20] The data is provided by the MoJ. In England and Wales, prisons are divided into categories (A,B,C and D) based on the severity of crime committed by inmates and the risk posed should the person escape. In order to obtain a sample that is sufficiently comparable, we focus on local male category B and C prisons. These categories hold prisoners who do not require maximum security, but who can not be trusted in open conditions.

[21] See Banker and Morey (1986) for a detailed analysis of including exogenous inputs and outputs in DEA.

Figure 3.3: Empirical prison production model



- We include in the model overall expenditures as joint input (tabulated in Table 3.1 per prisoner). The overall expenditures include security and rehabilitation costs, as well as accommodation and infrastructure related costs such as buildings and Information Technology, and food and utilities. The expenditures are deflated to allow for comparison over time.[22] Note that the overall expenditures include outside-prison expenditures of collaborating agencies. We thus fully acknowledge that prisons in England and Wales are not stand alone institutions. They closely collaborate with institutions that strife for improving reintegration and reducing reoffending risk (i.e., MAPPA, the Probation Service and Primary Care Trusts). Furthermore, a substantial part of the expenditures is payroll. Prison officers traditionally keep inmates secure and maintain order. However, prison officers also promote anti-bullying and suicide prevention policies, take part in programmes to help prisoners reflect on their offending behavior and prepare inmates for release through rehabilitation programmes. We therefore consider the expenditures by prison as a joint input, which simultaneously contributes to the production of all outputs.

  Nevertheless, there is room for refining the analysis and increasing the discriminatory power of the proposed model by allocating resources to outputs. In principle, the

[22] We used the GDP deflator at market prices for financial years as provided by the HM Treasury and use restated versions of the overall expenditures that improve comparability over time.

methodology allows to include both joint and output-specific expenses in the analysis. For this to be possible, the MoJ could construct a data set – which is consistent over time and over prisons – that breaks down the overall prison costs. For particular costs, allocation to outputs is straightforward. For example, the costs of food provision (which is frequently outsourced) can be allocated to the incarceration output. Costs of lecturers, workshop places, material and leaders can be attributed to the provision of activities. Administrative costs related to the discharges and re-integration into society could be attributed to the particular output on successful re-integration. Further, as many prisons collaborate with outside-prison organizations to provide out-of-cell activities and re-integration programs, the contractual agreements can be used to allocate resources to the outputs concerning purposeful activities and successful re-integration. However, for a substantial part of costs, allocation is less straightforward. To structure the input allocation to the multiple outputs, we advocate the use of '*activity based costing*' (ABC, see Cooper and Kaplan (1988)). The distinguishing feature of ABC is that costs are first attributed to activities and subsequently, these activity costs are allocated to the outputs. In comparison to other costing methodologies, which often are based on the produced output quantities, ABC gives a much clearer and more accurate picture of the production model of the multi-output decision making unit, see Cherchye et al. (2013). As such, ABC offers a framework to allocate expenses to particular outputs.

- To approximate the daily operations and administrative work that are needed to keep convicts outside society, we consider the places the prison offers to incarcerate prisons. In particular, we consider the Certified Normal Accommodation (CNA). By the Prison Act 1952, confining prisoners is only allowed in accommodation which is certified by an inspector that considers among others size, lighting, heating, communication-possibilities. Certified Normal Accommodation reflects the number of places the prison should not exceed (Prison Rules, 1999; rule 26). The respective cells are available for immediate use. Damaged cells, cells affected by building works and cells taken out of use due to staff shortages are excluded.[23]

[23]The study of the discretionary and non-discretionary aspects of overcrowding goes beyond this chapter. Results

- We simplify the provision of a humane prison environment and preparation for reintegration to organizing purposeful and outside-cell activities. The average prison population is included as non-discretionary output (which is non-discretionary to prison management but can be rescaled by higher-level decision makers) to control for the quantity of inmates for which purposeful and outside-cell activities can be provided. We include the total hours of purposeful activity per week[24] as an indicator of the effort during imprisonment that is taken to '*break the cycle*' by getting the prisoners to work and train outside their cell (see Ministry of Justice (2011)). In addition, we include time outside-cell[25] to fully acknowledge the beneficial aspects of other outside-cell activities such as sports and recreation.

- Successful reintegration is proxied by focusing on employment at discharge. Employment at release is a direct indicator of successful reintegration. Promoting employment at release is challenging. At most, 44 percent of prisoners have employment at release date. The number of discharges is included as non-discretionary output to control for the quantity of offenders that are released. Local prisons hold offenders with short sentences resulting in more discharges than the yearly average prison population. In 2009/10, the number of discharges was on average 1,310 while the average prison population in the sample was 880. In 2011/12, this average number increased to 1,526 while the number of average prison population held foot at 876. Successful reintegration also highly depends on the socio-economic environment in which prisoners are reintegrated. As this study deals with local prisons, we include the regional male unemployment rate[26] as output-specific environmental variable. We assume that the higher the regional male unemployment rate, the more difficult it is to find a job and thus to have employment at discharge.

- The three aspects of prison operation are highly conditioned by the heterogeneity in

available upon request show that however that inclusion of the overcrowding rate as environmental variable does not alter the results.

[24]This is calculated as the average number of hours purposeful activity per prisoner per week times the average prison population.

[25]This is calculated as 24 minus the average time within cell per prisoner, times the average prison population.

[26]The unemployment rates were retrieved from the Labour Force Survey (LFS), which is the largest household survey in the UK and provides the official measures of unemployment.

prisoner inflow. To take heterogeneity into account, we include the '*predicted rate of reoffending*' in an establishment as an environmental variable. The probability of reoffending is estimated at prison level by the Ministry of Justice (2011), based on prisoner-level data on social background, ethnicity, crime type, etc. We consider prisons with an inflow of prisoners which are more likely to re-offend as having a more difficult operating environment.[27] Note that it is possible to test for the direction of the environmental effect by following an approach suggested by Daraio and Simar (2005, 2007a) (see Appendix 3.D). We assume that the inflow of prisoners is beyond control of the prison management.

Only considering observations with the same z can lead to a dramatic loss of comparison partners. To overcome this issue we assume monotonicity in the sense that DMUs that can produce a given output in a more disadvantageous environment can also produce the same output in a similar environment as the DMU in question. As such, DMUs that operate in a less or equally advantageous environment can be used to construct the output-specific output sets.[28]

[27] It is sometimes argued that the quality of the socio-economic environment represents an input that is non-discretionary for the management. In fact, the environmental variables in our approach could be considered as undesirable inputs, which can not be scaled up or down. See Olesen et al. (2015) for a reference on how in general to include ratio measures, which are often used as contextual variables, as input and output data in DEA models.

[28] We follow Ruggiero (1996) to include environmental variables in the analysis, by including axiom 3.5 in Section 3.2.3.

	Year	Mean	St.Dev.	0%	25%	50%	75%	100%
Prison inputs								
Deflated overall resource expenditure per prisoner	2009/10	36601.76	7321.32	27248.38	32391.31	34962.52	38092.25	63691.60
(overallres)	2010/11	36697.52	8020.31	24938.20	32185.38	34627.37	39343.51	65503.96
	2011/12	34556.44	7150.36	23736.67	30400.02	32253.57	36836.95	59392.22
Daily operations and administrative work related to incarceration								
Certified Normal Accommodation	2009/10	666.21	291.94	145.00	449.25	646.67	872.75	1186.00
(CNA)	2010/11	688.76	299.74	146.00	464.50	682.00	938.00	1187.00
	2011/12	679.00	299.86	162.00	466.00	642.50	906.00	1187.00
Purposeful and outside-cell activities								
Average prison population	2009/10	879.63	351.71	232.92	638.02	843.92	1165.77	1653.58
(avpop)	2010/11	881.38	349.07	228.00	621.00	891.50	1170.25	1621.00
	2011/12	875.53	336.36	223.00	660.75	845.00	1121.75	1544.00
Average hours purposeful activity per week per prisoner	2009/10	20.63	3.72	16.29	18.09	19.91	22.29	34.98
(avpurp)	2010/11	21.32	3.83	16.90	18.68	20.48	23.30	35.20
	2011/12	20.99	3.61	16.81	18.52	20.05	22.11	33.73
Average hours outside-cell per day per prisoner	2009/10	8.16	1.30	5.60	7.30	7.90	9.20	12.10
(outcell)	2010/11	8.40	1.30	5.80	7.80	8.10	9.00	12.90
	2011/12	8.46	1.37	5.50	7.80	8.15	9.10	12.40
Successful reintegration of discharged offenders								
Number of Discharges	2009/10	1309.71	558.38	419.00	898.50	1207.00	1593.00	2933.00
(discharges)	2010/11	1417.60	519.15	460.00	1086.38	1376.50	1771.50	2575.50
	2011/12	1525.50	576.55	417.00	1169.25	1421.00	1973.50	2839.00
Percentage of discharges with employment	2009/10	24.89	6.68	12.80	19.93	23.60	28.80	41.70
(emprate)	2010/11	27.68	7.37	14.00	22.00	28.00	31.00	44.00
	2011/12	27.18	7.10	14.00	22.25	27.00	31.00	44.00
Prison characteristics								
Predicted rate of reoffending	2009/10	62.76	2.84	54.16	61.61	63.00	65.05	67.17
(predreof)	2010/11	62.54	5.52	51.00	58.88	62.27	67.33	77.28
	2011/12	62.09	4.76	53.40	58.87	61.20	65.91	74.18
Regional characteristics								
Regional male unemployment rate	2009/10	8.61	1.53	6.30	7.10	9.10	9.80	11.20
(regunemp)	2010/11	9.39	1.85	6.70	7.70	10.10	10.50	13.00
	2011/12	8.19	1.42	6.20	6.60	8.30	9.40	10.30

Table 3.1: Descriptive statistics

	overallres	cna	avpop	avpurp	outcell	discharges	emprate	regunemp	predreof
overallres	1.00	0.83	0.84	0.27	0.26	0.59	0.14	0.45	-0.40
cna	0.83	1.00	0.94	0.14	0.14	0.72	-0.07	0.47	-0.40
avpop	0.84	0.94	1.00	0.23	0.22	0.77	-0.12	0.54	-0.30
avpurp	0.27	0.14	0.23	1.00	0.71	0.39	0.29	-0.00	-0.05
outcell	0.26	0.14	0.22	0.71	1.00	0.36	0.32	0.06	-0.11
discharges	0.59	0.72	0.77	0.39	0.36	1.00	-0.01	0.44	-0.31
emprate	0.14	-0.07	-0.12	0.29	0.32	-0.01	1.00	-0.31	-0.37
regunemp	0.45	0.47	0.54	-0.00	0.06	0.44	-0.31	1.00	0.09
predreof	-0.40	-0.40	-0.30	-0.05	-0.11	-0.31	-0.37	0.09	1.00

Table 3.2: Correlogram

Table 3.2 shows the correlation between outputs, input and environmental variables. Output variables that relate to qualitative aspects of prison production are scaled per prisoner as in Table 3.1. There is modest correlation between input and outputs (even if not scaled per prisoner), indicating there can be deviations from optimal conduct or effects from the heterogeneity in the operating environment. Overall resources are positively associated with higher numbers of purposeful activity per prisoner (avpurp) and with time outside-cell (outcell). The correlation between input and the predicted rate of reoffending, is -0.4, indicating that prison characteristics are related to the size of the prison. Conditioning on the operating environment is thus needed to meaningfully analyze returns to scale.
As expected, the regional unemployment rate is negatively related to employment at discharge. Overall, Table 3.2 indicates a single output analysis cannot capture the production process of prisons as it would imply an omitted variable bias. We need an empirical analysis that includes multi-output structure and environmental heterogeneity to meaningfully estimate output-specific returns to scale in prison production.

3.4 Results

The methodology allows for a simultaneous analysis of output-specific scale efficiency and technical efficiency. We first discuss the results on scale efficiency in Subsection 3.4.1 and subsequently turn to technical efficiency in Subsection 3.4.2. As discussed in the methodology section, the order-m subsample bootstrapping routine (with m=50 and 1000 random draws)[29] is applied to lower the sensitivity of the efficiency estimates to potential outliers and extreme noise. In particular, the order-m efficiency measure is computed as the average over all draws. The order-m efficiency score can be interpreted as the expected efficiency score relative to a subsample of $m = 50$ prisons. Furthermore, this procedure allows us to report the statistical significance of the returns-to-scale estimations, which is based of the percentage of draws that leads to a particular returns to scale estimate. The most frequently estimated returns to scale is the one we report.

3.4.1 Output-specific returns to scale and scale efficiency

By applying the advocated framework on the empirical prison production model, we can examine whether output-specific returns to scale estimates differ considerably over outputs, implying a prison size dilemma for the public policy maker.

Incarceration Figure 3.4 shows returns to scale estimates and scale efficiency when the focus is solely on incarceration. The higher the values above 1, the more room for improvement. Scale efficiency estimates for the small and large prisons have values surpassing 1.5, indicating potential efficiency gains of over 50% by rescaling these prisons. These results are not surprising, since the cost per place varies between 31200 and 65500 pounds per year.

Given that the estimates are conditional upon the environment, it is possible for example that a particular prison is characterized by increasing returns to scale and that an even smaller prison is characterized by decreasing returns to scale. The reason is that the former prison is situated in a harsher environment than the latter. Although there is overlap between the returns to scale estimates, Figure 3.4 shows a clear pattern. Smaller prisons are generally characterized by increasing returns to scale and larger prisons are characterized by decreasing returns to scale.

Over the three book years, we find that 33 observations are characterized with decreasing returns to scale and 62 with increasing returns to scale. For respectively 25 and 56 obser-

[29] The value of m is chosen on the basis of visually inspecting the relation between the proportion of observations with $\varphi^m(v) < 1$ and the value of m as in Daraio and Simar (2007a).

vations the returns to scale estimates are significant at the 95% confidence level (see Table 3.4, 3.5 and 3.6). We find 7 prisons characterized by constant returns to scale, whereof 4 significant at the 95% confidence level. The returns to scale estimates show no year- specific patterns.[30]

The optimal scale size in terms of resources is around 30 to 40 million pounds. In terms of Certified Normal Accommodation, the prisons characterized by constant returns to scale provide between 606 and 1187 places. We therefore conclude that the optimal scale size of a prison with respect to incarceration is medium to large scale, depending on the environment. For the smallest prisons we do not find any prison that is scale efficient.

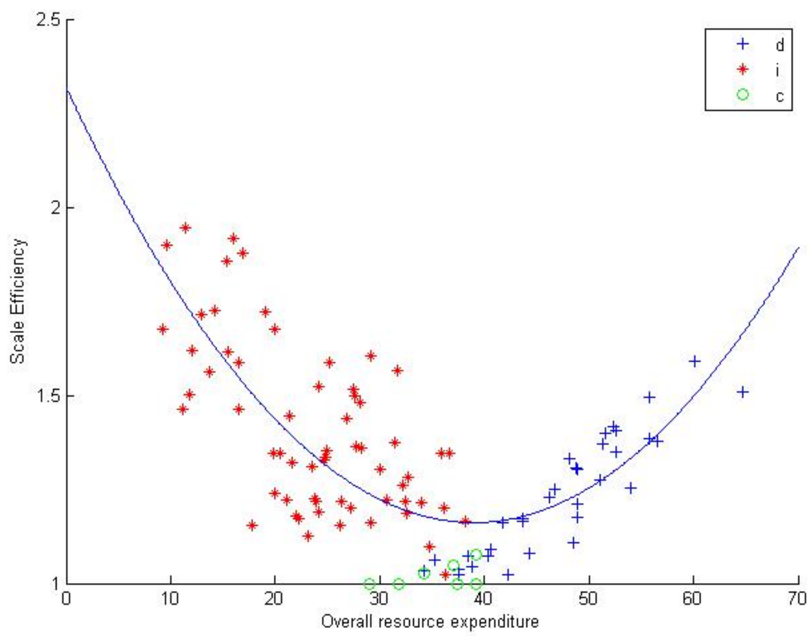


Figure 3.4: Returns to scale and scale efficiency with respect to incarceration, in function of overall resource expenditure.

Purposeful and outside-cell activities Figure 3.5 shows the estimates for the models that include the output variables that proxy purposeful and outside-cell activities to promote

[30]We did not find sub-constant returns to scale in our sample.

humane incarceration and to prepare inmates for reintegration into society. With this focus, we characterize 33 observations with *drs*, 63 with *irs* and 6 with *crs*. For respectively 22, 51 and 4 observations this is significant at the 95% confidence level. Overall, we find a similar pattern of returns to scale and scale efficiency as for the model focusing on keeping prisoners outside society. Focusing on purposeful and outside-cell activities, the prisons characterized by constant returns to scale provide between 606 and 1073 places. In terms of resources, the optimal scale size to provide activities is a little more spread than before. Still, we do not find supportive evidence for smaller prisons to be more scale efficient in terms of providing purposeful and outside-cell activities.[31]

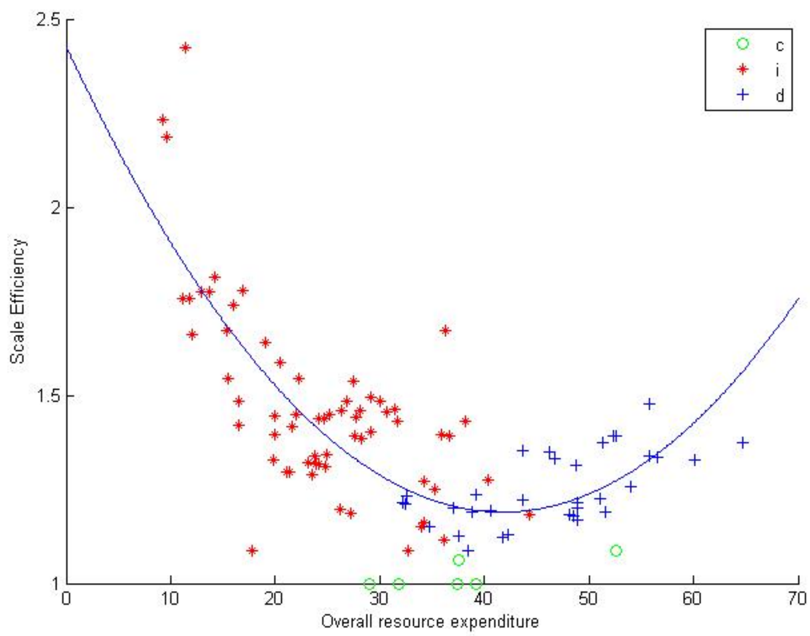


Figure 3.5: Returns to scale and scale efficiency with respect to purposeful and outside-cell activities, in function of overall resource expenditure.

Successful reintegration For successful reintegration we characterize respectively 32, 48 and 22 observations with respectively *drs*, *irs* and *crs*. For respectively 21, 22 and 18 ob-

[31] This result is robust for taking only one of the two proxies into the analysis.

servations the *rts* estimates are significant. Figure 3.6 shows a pattern of returns to scale and scale efficiency which highly differs in terms of optimal scale size over the considered environmental variables. The most productive scale size is in the broad interval of 10 to 50 million pounds and corresponds to prisons with a number of places between 221 and 1109.[32] Although smaller prisons can be optimal to provide reintegration, we find just as well medium and large prisons with an optimal scale size. Most probably, successful reintegration is highly dependent on unobserved heterogeneity in the effectiveness of reintegration programmes which is given the sample of observations difficult to disentangle from economies of scale.

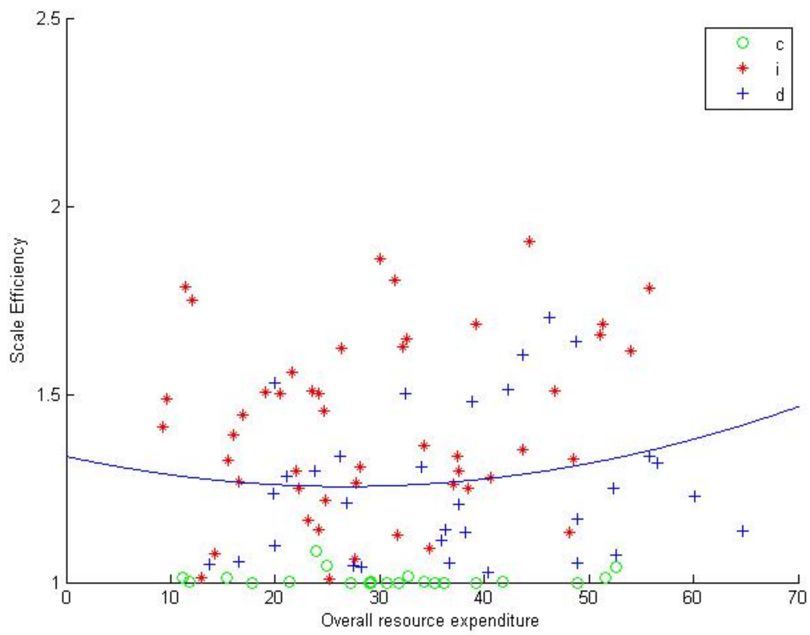


Figure 3.6: Returns to scale and scale efficiency with respect to successful reintegration, in function of overall resource expenditure.

[32]The high number of prisons exhibiting constant returns to scale is a consequence of the fact that we include the regional male unemployment rate as an additional environmental variable for the output successful reintegration. As a robustness check, we repeated the analysis by only controlling for prisoner characteristics, but not for regional characteristics. This results into only six prisons that exhibit constant returns to scale, however the technical and scale efficiency of some prisons is unrealistically high. Still, the optimal scale sizes remain equally scattered and no conclusions can be drawn about the optimal scale size with respect to successful reintegration.

In sum, our results over the three considered outputs reject the idea that public managers are faced with a prison size dilemma, which implies a choice between cost-per-place performance and social performance. We cannot reject medium to large scale to be optimal. Of course, the optimal scale size depends on the operating environment. For both incarceration and providing purposeful and outside-cell activities, we find supportive evidence that a medium to large scale size is optimal. For successful reintegration, we find no supportive evidence for drastic productivity gains by moving towards a very small prison scale.

Prison-specific estimates The overall finding that there is no prison size dilemma requires further prison-level consideration. Table 3.4, 3.5 and 3.6 in appendix shows that for 12 observations, we do find a prison size dilemma in the sense we estimate a prison with output-specific technologies to be simultaneously characterized by drs and irs, both significant at the 95% confidence level. For example, for HMP Bristol in book years 2009/10 and 2011/12, we find it is optimal to scale down prison scale to improve successful reintegration and scale up prison scale to improve the provision of purposeful and outside-cell activities and keeping convicts outside society by incarceration. From an operational viewpoint, allowing for output-specific and environment-specific returns to scale can be a valuable tool to provide policy advice on re-scaling prison conduct, taking the complexity of multi-output production into account.

3.4.2 Technical efficiency

Table 3.3 shows robust order-m technical efficiency estimates for the variable returns to scale model ($\varphi^m(v)$) and the constant returns to scale model ($\varphi^m(c)$). Order-m efficiency is reached when $\varphi^m = 1$. Note that the prison under evaluation is not necessarily included in the randomly drawn subsamples. Consequently, the order-m efficiency scores might be smaller than one. If the efficiency score is smaller than one, a prison is called super efficient. Overall, values of φ^m larger than one indicate that, on the basis of a subsample of m prisons, we estimate that there is room to proportionally increase the production, given the input and environment.

The first two columns show the respective vrs and crs results for the three outputs analyzed simultaneously, but allowing for output-specific technologies. The other columns show the results for the models that include only one output next to the input and output-specific environment.

With the exception of successful reintegration, Table 3.3 shows the technical efficiency of local male prisons in England and Wales is improving over time. Considering the multi-output model, on average, the room for increasing production went from 2 percent in 2009/10 to less than 0 percent (thus indicating super efficiency) in 2011/12. Stated differently, our estimates support the idea that the public policy of the coalition at place since 2010 was, at least partly, successful in reducing inefficiencies.[33] Still, some prisons considerably and persistently underperform (see Table 3.4, 3.5 and 3.6 in appendix). For example, technical efficiency of HMP Belmarsh is estimated to be respectively 1.36, 1.45 and 1.37 in the three consecutive book years considered. For HMP Manchester this is respectively 1.10,1.18 and 1.17.

In sum, using the advocated framework to consider multi-output prison production, we are able to pinpoint low performers in terms of both scale efficiency and technical efficiency. The persistent low performers require further attention from public managers. Are there additional prison characteristics that could explain the low figures?

	Multi-Output		Incarceration		Activities		Reintegration	
Year	$\varphi(v)$	$\varphi(c)$	$\varphi^1(v)$	$\varphi^2(c)$	$\varphi^2(v)$	$\varphi^2(c)$	$\varphi^3(v)$	$\varphi^3(c)$
2009/10	1,02	1,30	1,09	1,48	1,15	1,59	1,34	1,78
	(0,12)	(0,29)	(0,15)	(0,31)	(0,48)	(0,68)	(0,26)	(0,39)
2010/11	1,00	1,19	1,07	1,41	1,14	1,54	1,08	1,34
	(0,11)	(0,31)	(0,18)	(0,34)	(0,21)	(0,47)	(0,28)	(0,45)
2011/12	0,97	1,19	1,04	1,36	1,09	1,48	1,21	1,55
	(0,11)	(0,26)	(0,17)	(0,29)	(0,68)	(0,89)	(0,26)	(0,41)

Table 3.3: Mean (standard deviation) of order-m efficiency scores

A common argument is that the age of the prison, rather than size alone is a determinant of the performance outcome. In our sample, 23 prisons have been opened before 1900 and

[33] By pooling the three book years to perform the analysis, we assume that the prisons in the three consecutive years operate under the same technology. We therefore assume that the cost savings simply forced the prisons to operate in a more efficient manner and did not lead to a shift in technology.

9 prisons have been opened after 1990. However, many of the oldest prisons have been renovated. The most recently built prisons are on average larger than the older establishments. To separate the age effect from size, we compared the new prisons only with the largest older prisons.[34] We performed Wilcoxon ranksum tests for the null hypothesis that new and old prisons achieve the same efficiency level. Remarkably, the results indicate that the new prisons are not significantly more efficient in providing incarceration, nor in providing purposeful and outside cell activities nor in reintegrating prisoners[35].

3.4.3 Sensitivity analysis

Sensitivity analysis shows that our results are robust to altering the specification of the prison production model. The main findings are robust to altering the proxies for the three outputs. For example, our main findings are robust to including information on the number of assaults (for which data reliability can differ across prisons) or to extending the model to include accommodation at release, which is a necessary condition of successful reintegration. Next, we have tested whether there might be a time lag on the prison production process. To be concrete, we linked the inputs of a particular year to the output of the following year. However this did not result into better efficiency scores. Results from the sensitivity analysis are available upon request.

3.5 Conclusion

There is little reason to expect public firms to operate on their optimal scale size in the absence of competitive pressure. For prisons in England and Wales, there is a widespread policy debate concerning whether very small housing blocks or very large, so called '*titan*' prisons are the solution to improve the productivity of prisons. The general belief is that small prisons can provide a safe and humane environment wherein prisoners can be well

[34] We only compared prisons with a certified normal accommodation of at least 660 places.

[35] In total, we performed 9 Wilcoxon ranksum tests, for each of the three outputs and for several efficiency measures. For the outputs incarceration, activities and reintegration, the p-value for the order-m efficiency score under the assumption of variable returns to scale is respectively 77%, 25% and 85%. The p-value for the order-m efficiency score under constant returns to scale is respectively 53%, 40% and 41%. Moreover, the p-value for the scale efficiency is respectively 36%, 71% and 32%.

prepared to reintegrate into society and large prisons are especially effective when the focus is on expenditures-per-prisoner. If indeed the case, public managers would face a prison size dilemma. However, it is unclear whether the observed data support the idea of a prison size dilemma, as the policy debate does not go beyond an anecdotal discussion at most supported with partial indicators of reintegration and costs.

We provide a thorough examination of economies of scale using a complete multi-output assessment that allows for interlinkages between the output of incarcerating convicts and qualitative outputs (i.e., purposeful and outside-cell activity, successful reintegration) and allows that economies of scale can differ between the different qualitative dimensions of production.

Although our focus is on economies of scale in prisons, it is worth to note that the advocated methodological framework is more generally applicable to multi-output public sector organizations and multi-output manufacturing plants. For example Duncombe and Yinger (1993) study economies of scale in different dimensions of public production, with an application to fire protection.[36]

With respect to prisons in England and Wales, we do not find supportive evidence for the idea that public managers are confronted with a prison size dilemma. We cannot reject medium to large scale to be optimal. Depending on the operating environment, we find that medium to large scale size is optimal for both incarceration and providing purposeful and outside-cell activities. For successful reintegration, the results are mixed, but we do not find indications for drastic productivity gains by moving towards a very small prison scale. Our results are therefore supportive for a policy oriented towards 'titan' prisons, which are operated as a number of semi-autonomous units sharing a common site and set of services. However, it is worth noting that our results are based on a given set of observations of prison production. The building of very small and very large prisons can coincide with the introduction of new technologies, making extrapolation from the observed set of prison production difficult. Further research is needed to examine the optimal scale size to introduce productivity enhancing technological innovations.

Furthermore, the pillar of the UK 2010 coalition concerning the reduction of technical

[36]Duncombe and Yinger (1993) use a parametric translog cost function framework to study economies of scale.

inefficiencies within-prison is estimated to be, at least partly, successful. The technical efficiency is improving over the considered period 2009/10-2011/12 with the exception of the output successful reintegration.

We demonstrate the value of the multi-output production framework to analyze and test for output-specific scale (dis-)economies, using publicly available local prison data provided by the Ministry of Justice. As such, we provide a framework to test the success of recent policies to lower average costs by changing the scale of prisons.

Still, the Ministry of Justice could further increase the discriminatory power of the methodology by applying the advocated methodology with detailed information on the allocation of expenses to outputs and obtain even more detailed insight into the multi-output production process of local male prisons.

3.A Proofs

Proof of proposition 3.1. We first verify that $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ satisfies Axioms 3.1 - 3.6. Axiom 3.4 follows from the definition of β_n^r and $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$. Since $\beta_n^r(\mathbf{X}_n^r, rts^r) = 1$ and $n \in C_n^r(rts^r)$, we have that $\mathbf{y}_n^r \in \hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$. Moreover, Axioms 3.1 and 3.3 follow directly from the construction of $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ as the convex-monotone hull of the scaled output vectors of the DMUs in the set $C_n^r(rts^r)$. To verify Axiom 3.2, suppose that $\mathbf{X}^r \leq \mathbf{X}^{r*}$. By definition of β_s^r, we have that $\beta_s^r(\mathbf{X}^r, rts^r) \leq \beta_s^r(\mathbf{X}^{r*}, rts^r)$ and consequently $C_n^r(rts^r) \subset C_n^{r*}(rts^r)$. This implies that $\hat{P}^r(\mathbf{X}^r, \mathbf{Z}^r, rts^r) \subseteq \hat{P}^r(\mathbf{X}^{r*}, \mathbf{Z}^r, rts^r)$. To verify Axiom 3.5, suppose that $\mathbf{Z}^{r*} \leq \mathbf{Z}^r$. If $\mathbf{Z}^r \leq \mathbf{Z}_s^r$, then also $\mathbf{Z}^{r*} \leq \mathbf{Z}_s^r$, which implies that $C_n^r(rts^r) \subset C_n^{r*}(rts^r)$. Consequently, $\hat{P}^r(\mathbf{X}^r, \mathbf{Z}^r, rts^r) \subseteq \hat{P}^r(\mathbf{X}^r, \mathbf{Z}^{r*}, rts^r)$. Lastly, Axiom 3.6, is satisfied since $\beta_s^r(k^r\mathbf{X}^r, rts^r) = k^r\beta_s^r(\mathbf{X}^r, rts^r)$ for all $k^r \in K(rts^r)$. Then $\mathbf{y}^r \in \hat{P}^r(\mathbf{X}^r, \mathbf{Z}^r, rts^r)$ implies that $k^r\mathbf{y}^r \in \hat{P}^r(k^r\mathbf{X}^r, \mathbf{Z}^r, rts^r)$ for $k^r \in K(rts^r)$. We conclude that $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ satisfies Axioms 3.1 - 3.6.

It remains to prove that for any $P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ that satisfies Axioms 3.1 to 3.6, we have that $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r) \subseteq P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$. Take any $\mathbf{y}^{r*} \in \hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$. We need to prove that $\mathbf{y}^{r*} \in P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$. By the definition of $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$, we have given that

$$\mathbf{y}^{r*} \leq \sum_{s \in C_n^r(rts^r)} \lambda_s^r \beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{y}_s^r$$

for some $\lambda_s^r \geq 0$ such that $\sum_{s \in C_n^r(rts^r)} \lambda_s^r = 1$.

We can now prove that $\mathbf{y}^{r*} \in P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ by using that $P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ satisfies Axioms 3.1 to 3.6. First, Axiom 3.4 implies that

$$\mathbf{y}_s^r \in P^r(\mathbf{X}_s^r, \mathbf{Z}_s^r, rts^r) \forall s.$$

Using Axiom 3.6 we have that

$$\beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{y}_s^r \in P^r(\beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{X}_s^r, \mathbf{Z}_s^r, rts^r).$$

By definition of β_s^r we have that $\beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{X}_s^r \leq \mathbf{X}_n^r$. Together with Axiom 3.2 this implies that

$$\beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{y}_s^r \in P^r(\mathbf{X}_n^r, \mathbf{Z}_s^r, rts^r) \quad \forall s.$$

Due to Axiom 5

$$\beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{y}_s^r \in P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r) \quad \text{if } \mathbf{Z}_n^r \leq \mathbf{Z}_s^r.$$

Combining the definition of $\hat{P}^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$ with Axiom 3.3, this results into

$$\mathbf{y}^{r*} \leq \sum_{s \in C_n^r(rts^n)} \lambda_s^r \beta_s^r(\mathbf{X}_n^r, rts^r)\mathbf{y}_s^r \in P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$$

for $\lambda_s^r \geq 0$ such that $\sum_{s \in C_n^r(rts^r)} \lambda_s^r = 1$. Finally, Axiom 3.1 implies that $\mathbf{y}^{r*} \in P^r(\mathbf{X}_n^r, \mathbf{Z}_n^r, rts^r)$.

□

3.B Prison-specific estimates

Year 2009/10	Multi Output		Incarceration				Activities				Reintegration			
Prison	$\varphi(v)$	$\varphi(c)$	$\varphi^1(v)$	$\varphi^1(c)$	SE^1	RTS	$\varphi^2(v)$	$\varphi^2(c)$	SE^2	RTS	$\varphi^3(v)$	$\varphi^3(c)$	SE^3	RTS
Altcourse	0,91	0,96	1,33	1,80	1,35	d***	0,91	0,99	1,09	c	0,92	0,96	1,04	c
Bedford	0,97	1,45	0,97	1,66	1,72	i***	0,98	1,60	1,64	i***	0,97	1,45	1,51	i+
Belmarsh	1,36	2,05	1,36	2,05	1,51	d***	2,17	2,98	1,37	d***	3,28	3,73	1,14	d***
Birmingham	0,98	1,00	0,98	1,15	1,17	d***	1,19	1,42	1,20	d	1,00	1,00	1,00	c***
Bristol	1,00	1,33	1,17	1,43	1,23	i**	1,00	1,33	1,34	i***	1,28	1,66	1,30	d*
Brixton	1,00	1,38	1,00	1,38	1,37	i**	1,35	1,97	1,46	i***	1,48	2,67	1,80	i
Bullingdon	1,06	1,14	1,06	1,14	1,08	c	1,06	1,31	1,23	d	1,43	2,41	1,69	i
Cardiff	0,99	1,41	0,99	1,54	1,57	i***	0,99	1,41	1,43	i**	1,28	1,44	1,13	i*
Chelmsford	0,99	1,08	0,99	1,35	1,36	i***	1,08	1,50	1,38	i***	1,04	1,08	1,04	d
Doncaster	0,94	1,00	0,98	1,18	1,20	i***	0,94	1,05	1,11	i*	1,00	1,00	1,00	c***
Dorchester	0,99	1,76	1,10	2,13	1,95	i***	1,01	2,44	2,42	i***	0,99	1,76	1,79	i***
Durham	1,20	1,62	1,20	1,62	1,34	i***	1,23	1,71	1,39	i***	1,91	2,13	1,11	d***
Exeter	0,97	1,35	1,02	1,71	1,67	i***	0,97	1,35	1,40	i***	1,12	1,71	1,53	d
Forest Bank	0,93	0,99	1,12	1,16	1,04	d	0,93	0,99	1,06	c	1,07	1,29	1,21	d+
Gloucester	1,00	1,20	1,00	1,56	1,56	i***	1,00	1,77	1,78	i***	1,14	1,20	1,05	d*
High Down	1,00	1,10	1,02	1,10	1,08	d	1,42	1,68	1,18	i	1,00	1,90	1,91	i
Holme House	0,98	1,15	0,98	1,15	1,17	i***	1,22	1,74	1,43	i***	1,52	1,72	1,13	d***
Hull	0,99	1,34	1,00	1,34	1,34	i***	0,99	1,38	1,39	i**	1,28	1,35	1,05	d***
Leeds	1,00	1,03	1,21	1,29	1,07	d	1,27	1,62	1,28	i*	1,00	1,03	1,03	d***
Leicester	0,94	1,01	1,07	1,84	1,72	i***	0,99	1,80	1,82	i***	0,94	1,01	1,07	i*
Lewes	0,91	1,11	0,91	1,11	1,22	i**	1,06	1,55	1,46	i***	1,30	2,11	1,62	i
Lincoln	0,93	1,36	1,09	1,44	1,33	i***	1,11	1,59	1,44	i***	0,93	1,36	1,46	i
Liverpool	0,97	1,00	0,97	1,00	1,02	d	1,20	1,36	1,13	d+	1,64	2,48	1,51	d
Manchester	1,10	1,49	1,10	1,53	1,38	d***	1,11	1,49	1,34	d**	1,42	2,52	1,78	i
Norwich	0,97	1,32	0,97	1,32	1,36	i***	1,03	1,49	1,44	i***	1,20	1,51	1,26	i
Nottingham	1,26	1,84	1,29	1,91	1,48	i***	1,26	1,84	1,46	i***	2,32	3,03	1,31	i+
Parc	1,05	1,29	1,27	1,62	1,27	d***	1,05	1,29	1,23	d*	1,38	2,28	1,66	i+
Pentonville	1,16	1,41	1,16	1,41	1,21	d***	1,60	1,87	1,17	d***	1,39	1,47	1,05	d***
Preston	1,00	1,00	1,09	1,75	1,60	i***	1,00	1,49	1,49	i***	1,00	1,00	1,00	c**
Swansea	0,94	1,35	0,97	1,82	1,88	i***	0,94	1,67	1,78	i***	0,94	1,35	1,44	i***
Wandsworth	0,99	1,37	0,99	1,37	1,38	d***	1,10	1,47	1,34	d***	1,51	1,99	1,32	d***
Winchester	0,96	1,26	0,99	1,33	1,33	i***	0,96	1,26	1,31	i**	1,40	1,71	1,22	i+
Woodhill	1,39	2,02	1,62	2,02	1,25	d***	1,77	2,23	1,26	d*	1,39	2,24	1,61	i
Wormwood Scrubs	0,99	1,10	0,99	1,10	1,11	d*	1,30	1,53	1,18	d*	2,11	2,81	1,33	i+
Mean	*1,02*	*1,30*	*1,09*	*1,48*	*1,36*		*1,15*	*1,59*	*1,39*		*1,34*	*1,78*	*1,33*	
(Standard Deviation)	*(0,12)*	*(0,29)*	*(0,15)*	*(0,31)*	*(0,24)*		*(0,26)*	*(0,39)*	*(0,26)*		*(0,48)*	*(0,68)*	*(0,28)*	

Note: $+$ indicates over 90% of the subsample bootstrap replications show the value, $*$ indicates 95%, $**$ indicates 99% and $***$ indicates 99.9%.

Table 3.4: Order-m efficiency scores and returns to scale estimates (2009/10)

Year 2010/11	Multi Output		Incarceration				Activities				Reintegration			
Prison	$\varphi(v)$	$\varphi(c)$	$\varphi^1(v)$	$\varphi^1(c)$	SE^1	RTS	$\varphi^2(v)$	$\varphi^2(c)$	SE^2	RTS	$\varphi^3(v)$	$\varphi^3(c)$	SE^3	RTS
Altcourse	0,92	0,99	1,37	1,92	1,40	d***	0,92	1,10	1,19	d*	0,98	0,99	1,01	c*
Bedford	0,87	1,10	0,95	1,51	1,59	i***	0,91	1,30	1,42	i***	0,87	1,10	1,27	i+
Belmarsh	1,45	2,17	1,45	2,17	1,49	d***	2,17	3,21	1,48	d***	1,63	2,18	1,34	d***
Birmingham	0,99	1,22	1,04	1,22	1,17	d***	1,08	1,32	1,22	d*	0,99	1,34	1,35	i
Bristol	1,00	1,00	1,00	1,44	1,44	i***	1,00	1,30	1,30	i***	1,00	1,00	1,00	c***
Brixton	1,00	1,57	1,20	1,57	1,30	i**	1,41	2,10	1,48	i***	1,00	1,86	1,86	i***
Bullingdon	0,95	1,16	0,95	1,16	1,22	i**	1,00	1,22	1,21	d	1,56	2,34	1,50	d
Cardiff	0,97	1,06	0,99	1,49	1,50	i***	0,97	1,34	1,39	i**	1,00	1,06	1,06	i***
Chelmsford	0,95	1,00	0,99	1,34	1,35	i***	0,97	1,30	1,34	i***	0,95	1,00	1,05	c
Doncaster	0,78	0,93	0,93	1,11	1,20	i*	0,78	0,93	1,18	i	1,00	1,00	1,00	c***
Dorchester	1,00	1,49	1,00	1,90	1,90	i***	1,00	2,18	2,18	i***	1,00	1,49	1,49	i***
Durham	1,00	1,00	1,00	1,00	1,00	c***	1,00	1,00	1,00	c***	1,00	1,00	1,00	c***
Exeter	1,00	1,00	1,00	1,15	1,15	i***	1,00	1,09	1,09	i***	1,00	1,00	1,00	c***
Forest Bank	1,00	1,00	1,00	1,00	1,00	c***	1,00	1,00	1,00	c***	1,00	1,00	1,00	c***
Gloucester	1,00	1,00	1,00	1,50	1,50	i***	1,00	1,76	1,76	i***	1,00	1,00	1,00	c**
High Down	0,99	1,09	0,99	1,09	1,10	i	1,37	1,57	1,15	d	1,00	1,09	1,09	i*
Holme House	0,98	1,00	0,98	1,00	1,02	i+	1,00	1,67	1,67	i**	1,07	1,22	1,14	d***
Hull	1,00	1,00	1,00	1,22	1,22	i***	1,00	1,45	1,45	i***	1,00	1,00	1,00	c***
Leeds	1,00	1,00	1,19	1,27	1,06	d	1,20	1,50	1,25	i+	1,00	1,00	1,00	c***
Leicester	0,99	1,00	1,07	1,83	1,72	i***	1,00	1,78	1,77	i***	0,99	1,00	1,01	i
Lewes	0,88	1,04	0,88	1,04	1,18	i*	1,29	1,87	1,45	i***	0,89	1,16	1,30	i
Lincoln	0,92	1,43	1,09	1,45	1,32	i***	1,21	1,71	1,42	i***	0,92	1,43	1,56	i*
Liverpool	0,95	0,99	0,96	0,99	1,02	d	1,22	1,38	1,13	d+	0,95	1,23	1,29	i
Manchester	1,18	1,62	1,22	1,67	1,37	d***	1,18	1,62	1,38	d*	1,67	2,83	1,69	i
Norwich	0,89	1,13	0,95	1,13	1,19	i**	0,91	1,19	1,32	i***	0,89	1,34	1,50	i*
Nottingham	1,05	1,21	1,18	1,43	1,22	i**	1,05	1,21	1,15	i*	1,22	1,59	1,31	d*
Parc	1,05	1,46	1,15	1,63	1,42	d***	1,05	1,46	1,39	d***	1,52	1,90	1,25	d***
Pentonville	1,00	1,00	1,17	1,36	1,16	d***	1,47	1,65	1,12	d***	1,00	1,00	1,00	c***
Preston	1,00	1,01	1,00	1,60	1,59	i***	1,00	1,45	1,45	i***	1,00	1,01	1,01	i***
Swansea	0,94	1,00	0,99	1,83	1,86	i***	0,94	1,57	1,67	i***	0,99	1,00	1,01	c*
Wandsworth	1,06	1,38	1,06	1,38	1,30	d***	1,33	1,62	1,22	d**	1,30	1,51	1,17	d***
Winchester	0,99	1,21	0,99	1,21	1,22	i***	0,99	1,29	1,30	i***	1,00	1,28	1,28	d*
Woodhill	1,26	2,14	1,79	2,20	1,23	d***	1,85	2,49	1,35	d*	1,26	2,14	1,70	d
Wormwood Scrubs	1,00	1,09	1,00	1,09	1,09	d*	1,44	1,71	1,19	d*	1,19	1,53	1,28	i*
Mean	*1,00*	*1,19*	*1,07*	*1,41*	*1,31*		*1,14*	*1,54*	*1,35*		*1,08*	*1,34*	*1,22*	
(Standard Deviation)	*(0,11)*	*(0,31)*	*(0,18)*	*(0,34)*	*(0,23)*		*(0,28)*	*(0,45)*	*(0,24)*		*(0,21)*	*(0,47)*	*(0,24)*	

Note: $+$ indicates over 90% of the subsample bootstrap replications show the value, $*$ indicates 95%, $**$ indicates 99% and $***$ indicates 99.9%.

Table 3.5: Order-m efficiency scores and returns to scale estimates (2010/11)

Year 2011/12	Multi Output		Incarceration				Activities				Reintegration			
Prison	$\varphi(v)$	$\varphi(c)$	$\varphi^1(v)$	$\varphi^1(c)$	SE^1	RTS	$\varphi^2(v)$	$\varphi^2(c)$	SE^2	RTS	$\varphi^3(v)$	$\varphi^3(c)$	SE^3	RTS
Altcourse	0,94	1,07	1,34	1,78	1,33	d***	0,94	1,11	1,18	d*	0,94	1,07	1,13	i
Bedford	0,85	1,13	0,87	1,41	1,62	i***	0,85	1,32	1,55	i***	0,86	1,13	1,32	i*
Belmarsh	1,37	1,92	1,37	1,92	1,41	d***	2,10	2,92	1,39	d***	2,82	3,03	1,07	d***
Birmingham	0,85	0,92	1,10	1,34	1,22	i*	1,08	1,43	1,32	i**	0,85	0,92	1,08	c
Bristol	0,99	1,22	0,99	1,33	1,34	i***	0,99	1,31	1,33	i***	0,99	1,22	1,23	d*
Brixton	1,04	1,50	1,04	1,50	1,44	i***	1,26	1,87	1,49	i***	1,56	1,89	1,21	d**
Bullingdon	0,91	1,14	0,91	1,14	1,26	i**	1,07	1,30	1,21	d	1,02	1,66	1,63	i
Cardiff	0,98	1,18	0,98	1,48	1,52	i***	1,00	1,53	1,54	i**	1,13	1,18	1,04	d+
Chelmsford	0,89	1,15	0,95	1,24	1,31	i***	0,89	1,15	1,29	i**	0,94	1,42	1,51	i
Doncaster	0,72	0,87	0,94	1,08	1,16	i	0,72	0,87	1,20	i	0,99	1,32	1,33	d*
Dorchester	0,98	1,39	0,98	1,64	1,68	i***	0,99	2,21	2,23	i***	0,98	1,39	1,41	i***
Durham	1,00	1,00	1,00	1,00	1,00	c***	1,00	1,00	1,00	c***	1,00	1,00	1,00	c***
Exeter	1,00	1,05	1,00	1,46	1,46	i***	1,00	1,48	1,48	i***	1,00	1,05	1,06	d*
Forest Bank	1,00	1,00	1,00	1,00	1,00	c***	1,00	1,00	1,00	c***	1,00	1,34	1,34	i***
Gloucester	0,98	0,99	0,99	1,45	1,46	i***	0,99	1,74	1,76	i***	0,98	0,99	1,01	c*
High Down	0,86	1,02	0,86	1,02	1,19	i	1,22	1,51	1,23	d	0,90	1,49	1,65	i**
Holme House	0,97	0,99	0,97	0,99	1,03	c	0,98	1,25	1,27	i**	1,20	1,64	1,36	i**
Hull	0,99	1,00	0,99	1,15	1,16	i**	0,99	1,39	1,40	i**	1,00	1,00	1,00	c***
Leeds	0,99	1,00	1,19	1,23	1,03	d	0,99	1,14	1,16	i+	1,00	1,00	1,00	c**
Leicester	0,90	1,57	1,06	1,72	1,62	i***	0,97	1,61	1,66	i***	0,90	1,57	1,75	i***
Lewes	0,86	1,04	0,89	1,04	1,17	i+	1,07	1,65	1,54	i***	0,86	1,08	1,25	i
Lincoln	0,92	1,38	1,03	1,39	1,34	i***	0,96	1,52	1,59	i***	0,92	1,38	1,50	i*
Liverpool	0,91	0,95	0,91	0,95	1,05	c	1,10	1,32	1,20	d	0,92	1,16	1,26	i*
Manchester	1,17	1,53	1,22	1,59	1,31	d***	1,17	1,53	1,31	d**	2,24	3,67	1,64	d
Norwich	0,90	1,09	0,97	1,09	1,13	i+	0,90	1,18	1,32	i***	0,96	1,12	1,17	i
Nottingham	0,97	0,99	1,05	1,35	1,28	i***	0,97	1,05	1,09	i+	0,98	0,99	1,01	c+
Parc	0,99	1,43	0,99	1,57	1,59	d***	1,08	1,43	1,33	d***	4,31	5,29	1,23	d***
Pentonville	0,87	1,08	1,15	1,23	1,07	d*	1,37	1,49	1,09	d	0,87	1,08	1,25	i
Preston	1,00	1,14	1,00	1,52	1,52	i***	1,00	1,44	1,44	i***	1,00	1,14	1,14	i***
Swansea	0,93	1,30	0,99	1,89	1,92	i***	0,94	1,63	1,74	i***	0,93	1,30	1,39	i*
Wandsworth	1,06	1,32	1,06	1,32	1,25	d***	1,18	1,57	1,33	d*	1,26	1,91	1,51	i+
Winchester	0,97	1,06	0,97	1,20	1,24	i***	0,99	1,43	1,45	i***	0,97	1,06	1,10	d
Woodhill	1,22	1,95	1,78	2,08	1,17	d***	1,76	2,38	1,35	d*	1,22	1,95	1,61	d
Wormwood Scrubs	0,99	1,04	0,99	1,04	1,04	d+	1,44	1,71	1,19	d+	1,60	2,37	1,48	d
Mean	*0,97*	*1,19*	*1,04*	*1,36*	*1,30*		*1,09*	*1,48*	*1,37*		*1,21*	*1,55*	*1,28*	
(Standard Deviation)	*(0,11)*	*(0,26)*	*(0,17)*	*(0,29)*	*(0,22)*		*(0,26)*	*(0,41)*	*(0,24)*		*(0,68)*	*(0,89)*	*(0,22)*	

Note: $+$ indicates over 90% of the subsample bootstrap replications show the value, $*$ indicates 95%, $**$ indicates 99% and $***$ indicates 99.9%.

Table 3.6: Order-m efficiency scores and returns to scale estimates (2011/12)

3.C Dual formulation

In this section we discuss the dual formulation of the linear programming problem (LP-1) in terms of revenue efficiency. We categorize each output as either discretionary, and assign it to an index set I_d^r, or fixed, and assign it to an index set I_f^r, such that $I_d^r \cup I_f^r = \{1, \ldots, M^r\}$. We obtain as dual problem (LP-2):

$$
\begin{aligned}
RE_n^r(rts^r) \quad &= \quad \min_{o \geq 0, P_m^r \geq 0} o - \sum_{m \in I_f^r} y_{mn}^r P_m^r \\
s.t. \quad & \\
\text{(R-1)} \qquad & o \geq \sum_{m \in I_d^r} P_m^r \beta_s^r(\mathbf{X}_n^r, rts^r) y_{ms}^r + \sum_{m \in I_f^r} P_m^r \beta_s^r(\mathbf{X}_n^r, rts^r) y_{ms}^r \quad \forall s \in C_n^r(rts^r) \\
\text{(R-2)} \qquad & \sum_{m \in I_d^r} y_{mn}^r P_m^r = 1.
\end{aligned}
$$

To explain the revenue efficiency interpretation, we need to interpret the variables P_m^r as shadow prices, which are used to value the outputs. In particular, P_m^r is the shadow price of output y_m^r for $m \in \{1, \ldots, M^r\}$. Next, the variable o can be interpreted as the maximal (total) revenue that can be achieved, conditional on the level of input and the environment of DMU n. In particular, constraint (R-1) specifies that the maximal revenue level can not be lower than the revenue level associated with any other DMU s which belongs to the comparison partners of DMU n ($s \in C_n^r(rts^r)$). Note that the total revenue consists of a discretionary part and a fixed part. Since a DMU has no control over the fixed outputs, only the discretionary outputs should be taken into account to determine the revenue efficiency. Constraint (R-2) defines a normalization of the discretionary revenue for the evaluated DMU n. Because DMU n's discretionary revenue level is normalized to 1, this makes that the objective function value $o - \sum_{m \in I_f^r} y_{mn}^r P_m^r$ can be interpreted as the ratio of the maximum discretionary revenue, over the DMU's actual discretionary revenue. The above presented linear programming problem then selects the shadow prices such that the ratio of the maximum discretionary revenue over the actual discretionary revenue is minimized.

3.D Robustness analysis environmental variables

We followed Ruggiero (1996) to control for environmental variables that affect production. Following this approach, we need information on the direction of the influence of the environmental factors. However, there are other possibilities to include environmental variables in the analysis, which do not need such information. For example, Daraio and Simar (2005, 2007a) follow a probabilistic approach to include environmental variables. The authors introduce a robust DEA estimator which conditions on external environmental factors, using a kernel weighting approach. With this approach, prisons in a similar environment as the prison under evaluation will have a larger probability to be drawn as a reference prison. In particular, the samples of size m are constructed by drawing with replacement among those $\mathbf{Y}_s^r$ such that $\mathbf{X}_s^r \leq \mathbf{X}^r$ and such that the probability of drawing a prison corresponds to

$$\frac{K((Z^r - Z_s^r)/h)}{\sum_{t=1}^{N} K((Z^r - Z_t^r)/h)},$$

where K denotes a Kernel function and h an appropriate bandwidth. An advantage of this method is that it is also possible to investigate the direction of the impact of the environmental variables on the production process. As a robustness check, we test in this section the direction of the environmental effect by using the approach suggested by Daraio and Simar (2005, 2007a). These papers suggest to plot the ratio of the conditioned to the unconditioned efficiency score, in function of the environmental affect. An increasing pattern reveals a favorable effect of the environmental variable, a decreasing pattern reveals an unfavorable effect. To compute the conditioned efficiency score, we have used an Epanechnikov kernel and estimated the bandwidth using the Sheater Jones rule of thumb. Since our analysis on the optimal scale size has shown that the prison production process does not exhibit constant returns to scale, we apply the approach of Daraio and Simar (2005, 2007a) for the variable returns-to-scale model. Note that we test the effect of one environmental variable at a time, however it is perfectly possible investigate the impact of a multivariate vector of environmental variables (see Daraio and Simar (2007a)).

First, we consider the effect of the predicted rate of reoffending on each of the three outputs. We find that the predicted rate of reoffending indeed has a negative impact on the outputs incarceration (see Figure 3.7) and successful reintegration (see Figure 3.8), which confirms our assumption that the predicted rate of reoffending has an unfavorable impact on the prison production process. However, the effect of predicted rate of reoffending on the provision of purposeful and outside cell activities is less clear. In particular, the effect of the predicted rate of reoffending seems to have an inverted u-shaped function, indicating that there is first a (small) favorable effect of predicted rate of reoffending and consequently a (small) unfavorable effect. In principle, we could repeat the analysis on the optimal scale size employing the approach suggested by Daraio and Simar (2005, 2007a) to control for predicted rate of reoffending. This method allows the predicted rate of reoffending to have a non-monotone effect on the provision of purposeful and outside cell activities. For ease of application for the practitioner, we have chosen to maintain the assumption of an unfavorable impact of predicted rate of reoffending, and employ Ruggiero (1996) to control for environmental variables. We argue that it might be interesting to take a closer look at three of the observations with the lowest predicted rate of reoffending. These three observations seem to be much less efficient if they are benchmarked against the whole sample. Since it is counter intuitive that the predicted rate of reoffending has a favorable effect, there might be other confounding factors which led to this result.

Further, we consider the effect of the regional male unemployment rate on the reintegration output. We find - as expected - that the regional male unemployment rate has an unfavorable impact on successful reintegration.

In principle, the age of the prison could also be seen as an environmental variable. We check with the approach of Daraio and Simar (2005, 2007a) what the impact is on the three outputs, when controlling for the construction year of the prison. In particular, we find that the most recently built prisons (after 1990) have an advantage for the output incarceration. We also find a positive effect of construction year for the output activities. Finally, for the output reintegration, we find almost no effect of the construction year of the prison. In principle, we could also have conditioned on the construction year of the prisons, to perform the analysis on the most productive scale size. However, if we would condition on an additional environmental variable, we would end up with a very small sample of comparison partners for most prisons (the curse of dimensionality). Remember that in Section 3.4.2, we did not find significant differences between the average efficiency of old versus new (large) prisons. This indicates that, once we condition on the predicted rate of reoffending and the regional male unemployment rate, we find little effect of the age of the prison.

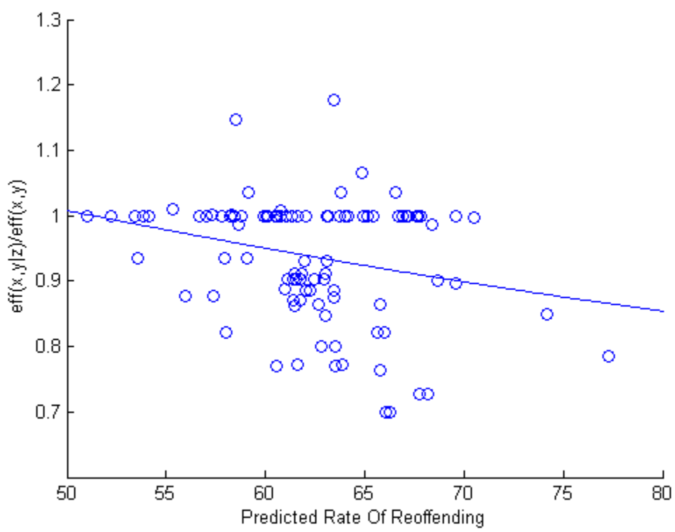


Figure 3.7: Effect of predicted rate of reoffending on efficiency ratio with respect to incarceration

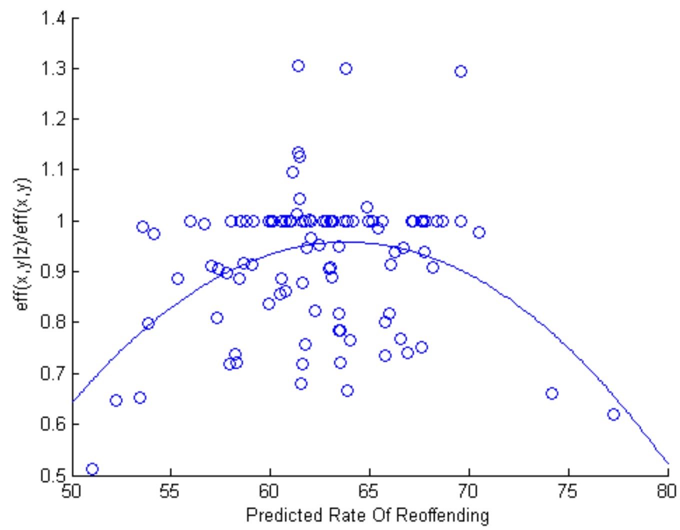


Figure 3.8: Effect of predicted rate of reoffending on efficiency ratio with respect to activities

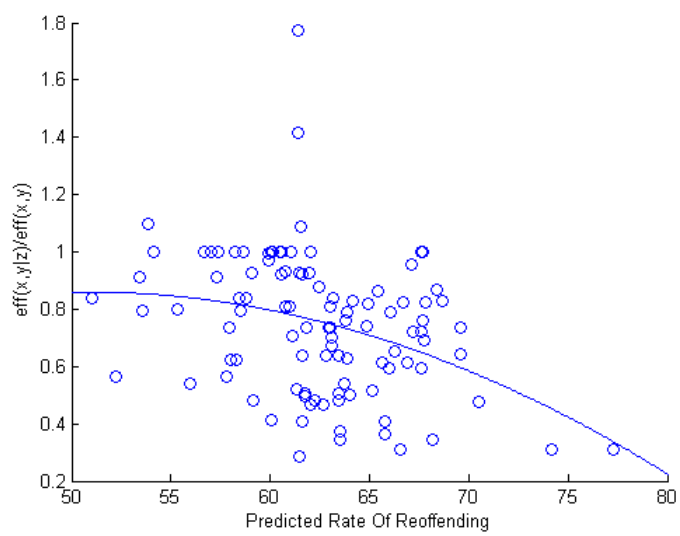


Figure 3.9: Effect of predicted rate of reoffending on efficiency ratio with respect to reintegration

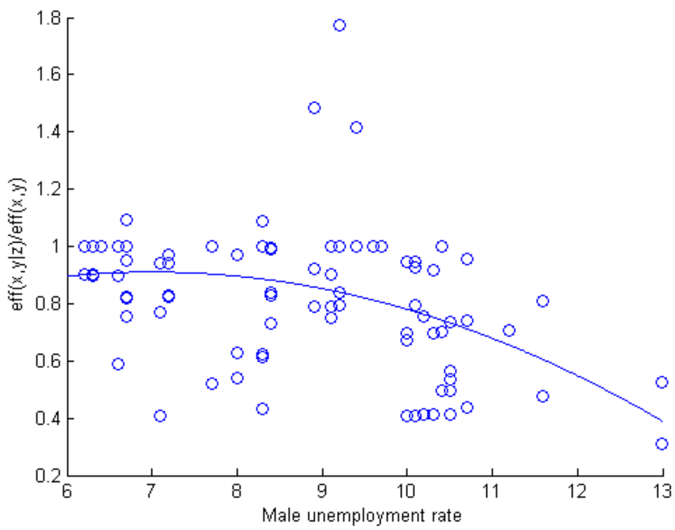


Figure 3.10: Effect of regional male unemployment rate on efficiency ratio with respect to reintegration

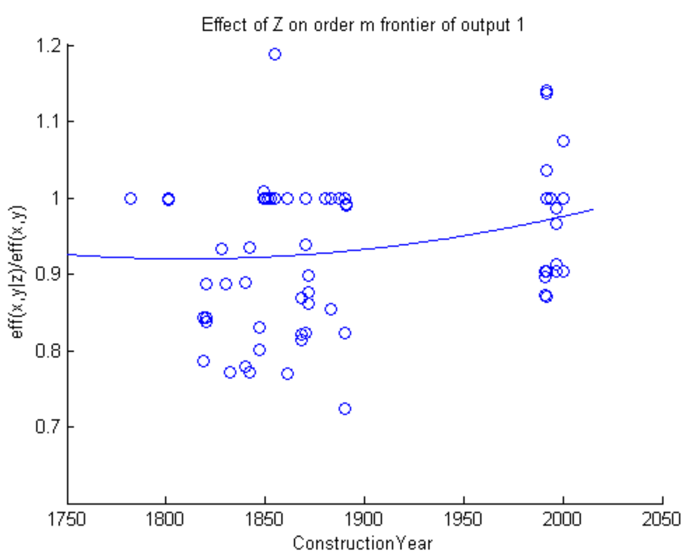


Figure 3.11: Effect of construction year on efficiency ratio with respect to incarceration

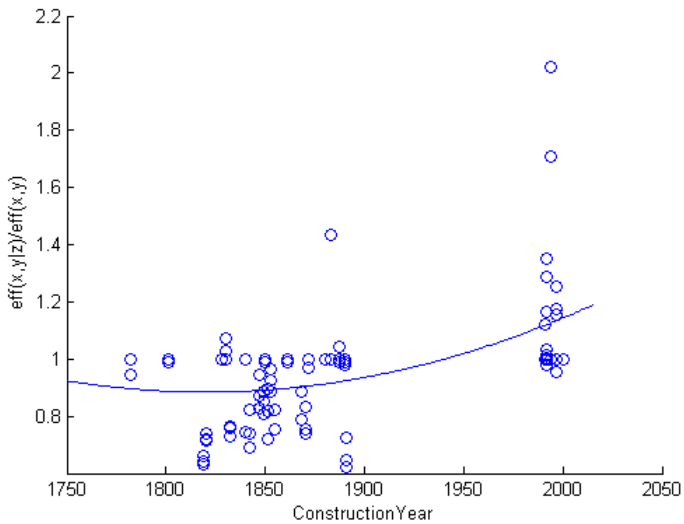


Figure 3.12: Effect of construction year on efficiency ratio with respect to activities

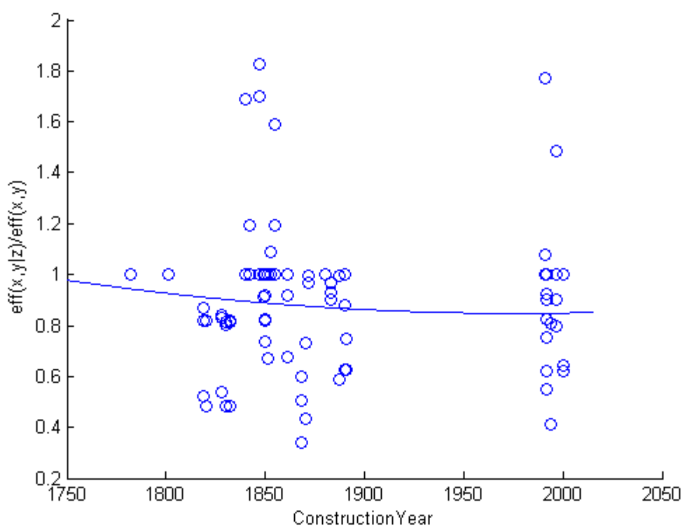


Figure 3.13: Effect of construction year on efficiency ratio with respect to reintegration

Part III

Consumption behavior

Chapter 4

Revealed preferences for time with children

Abstract

We study a labor supply model with home production of child well-being. In this model, child well-being is produced by means of parental time and expenditures. The distinguishing feature of our study is that we allow for process benefits. We take into account that part of the parent's time spent with children is also a substitute for the caring parent's leisure. Our approach to modeling process benefits is novel in the sense that it does not require a parametric specification of the household's utility function and of the household's production technology. Moreover, the output of the home production need not be observed. This makes our test particularly useful to investigate nonmarketable goods such as child well-being. We apply our model to a data set on consumption and time use choices of Dutch households. We find that process benefits significantly improve the goodness-of-fit of our labor supply models. We also recover the distribution of process benefits in the sample. For robustness, we apply our methodology both to a unitary and a collective labor supply model.[1]

[1] This chapter is based on joint work with Sam Cosaert (KU Leuven). We thank Bart Capéau, Laurens Cherchye, Pierre-André Chiappori, Bram De Rock, François Maniquet and Frederic Vermeulen for many helpful comments. We are also grateful for the referee reports from the reading group on collective models of household consumption at ULB. Our thanks also goes to the members of the research center of Public Economics Leuven and participants of the EEA conference in Mannheim. Finally, the LISS panel data were collected by CentERdata (Tilburg University, The Netherlands) through its MESS project funded by the Netherlands Organization for Scientific Research

4.1 Introduction

Spending time in more or less enjoyable ways has an impact on our well-being. There is no doubt that leisure-activities, such as relaxing and socializing, are enjoyable. However many people also derive pleasure from non-leisure activities. Juster (1985) designed a set of questions to investigate the extent to which various activities are enjoyable. In particular, Juster (1985) asked respondents in a US time survey, to rate how much they generally enjoy a given type of activity, such as cleaning the house, cooking, taking care of the children and so forth. Juster (1985) specifically asked the respondents to keep in mind that he was interested in whether they liked the activity, irrespective of whether they think it is important to do. To capture the utility that accrues during particular activities, Dow and Juster (1985) and Juster et al. (1985) introduce the notion of process benefits. Similarly, Hallberg and Klevmarken (2003) used Swedish data to rank activities on the basis of the reported process benefits. The authors found that households perceive spending time with their children as one of the most enjoyable activities. It is more enjoyable than watching TV, market work and reading books or magazines, on average. Schwarz et al. (2009) and Krueger et al. (2009) argue that asking people about how they feel about activities in general provides a different ranking than when their actual experiences are used to guide their reported feelings during those activities. An alternative to the set of questions designed by Juster (1985), is a time use diary to measure process benefits (Gershuny and Halpin (1996), Robinson and Godbey (1997) and Krueger et al. (2009)). For example, Krueger et al. (2009) work with a time use diary in which individuals can indicate the nature of the activity and the extent to which various emotions (happy, tired, stressed, sad, interested and pain) were present or absent. Still, Krueger et al. (2009) also find that child care belongs to the more enjoyable activities, more than housework and working.

The above mentioned studies attempt to directly measure process benefits. However, we will follow Graham and Green (1984), Kerkhofs and Kooreman (2003) and Gørtz (2011) and apply a structural labor supply model to analyze process benefits, on the basis of consumption and time use data. Doing so, there is no need to directly observe the process benefits. Instead, we test whether models with varying degrees of process benefits can describe the observed behavior.

We focus in particular on process benefits related to child care. We explicitly model the fact that parents may enjoy caring for the children by assuming that child care time is to some extent a substitute for leisure. In line with existing models, parents of course also invest time in their children because they care about the well-being of the children. Child care time therefore 'jointly' produces child well-being and leisure for the caring parent. The challenge is that child well-being is neither observed nor marketable. For this reason, we follow a nonparametric - revealed preference - approach which requires no information on the output or price associated with child well-being. Moreover, our approach is independent of functional restrictions on the parents' preferences.

Literature The theory of time allocation, initiated by Becker (1965) and Gronau (1977), typically assumes that household members allocate their time to either market work, leisure or some domestic production activity. Indeed, households can produce certain 'domestic' commodities by allocating their time and resources to the corresponding activity. Notable examples include the cleaning of the house, cooking, or caring for the household's own children. In the latter case, child well-being is treated as a domestic commodity which is produced by a combination of time spent with children and consumption goods allocated to children.

Remarkably, the models by Becker (1965) and Gronau (1977) do not allow joint production. However, it seems reasonable that time allocated to the production of a commodity also 'directly' impacts on utility, when household members enjoy spending time in that particular way. Pollak and Wachter (1975) even argue that when time is an input in the household production process, joint production is the rule and not an exception.[2] Graham and Green (1984), Kerkhofs and Kooreman (2003) and Gørtz (2011) extended Gronau (1977)'s model by implementing process benefits. They allow time spent on a particular commodity to generate both *indirect* utility, through the commodity which is produced, and *direct* utility in

[2]Pollak and Wachter (1975) provide a critique on the household production function approach. They argue that only in the absence of joint production and under the assumption of constant returns to scale, the demand for commodities should be analyzed in terms of commodity prices. If not, the implicit commodity prices depend on the commodity bundle consumed. At the same time, the authors argue that joint production is a realistic scenario. Although allowing for joint production thus complicates the estimation of the demand for commodities, it is still perfectly possible to analyze the allocation of goods and time among household activities.

terms of leisure. Specifically, the authors treat some part of the time input as a substitute to pure leisure time.

In this chapter, we present a revealed preference test for rational consumption and time use that takes into account that parents enjoy process benefits related to child care. This methodology enables us to identify process benefits for mothers and fathers. The contribution of this chapter to the literature is three-fold.

Contributions First, we propose an empirical test to verify whether we can explain the observed child care behavior of parents by a household labor supply model with home production and process benefits. In contrast to Graham and Green (1984), Kerkhofs and Kooreman (2003) and Gørtz (2011) who focus on *estimating* a parametric version of the model with and without process benefits, this chapter aims at *testing* whether models with varying degrees of process benefits can describe the observed behavior. Given this focus, we develop a characterization in which neither the household's objective function nor the intra-household production technology are specified. This nonparametric approach avoids a so called 'dual hypothesis'. Indeed, in a parametric framework, a rejection of rationality might indicate irrational behavior on behalf of the household, but also an incorrect specification of the household's utility function and/or production function. By following a revealed preference approach, we can rule out the issues corresponding to the dual hypothesis. Although we do not impose a functional specification for the utility function nor for the household production function, we will do so for the process benefits. In particular, we follow Graham and Green (1984), Kerkhofs and Kooreman (2003) and Gørtz (2011) for a functional specification of the process benefits which is fully determined by a single parameter δ^i. This parameter δ^i will allow us to quantify the degree of process benefits in a straightforward manner.

The revealed preference approach enables us to study smaller samples and allows us to vary the process benefits across individuals. This gives insight in the distribution of process benefits across the sample. We build on the revealed preference characterizations by Samuelson (1938), Afriat (1967), Diewert (1973) and Varian (1982). The revealed preference conditions in the context of home production are closely related to revealed preference conditions that characterize data sets that are consistent with the maximization of a weakly separable utility function (Varian (1983) and Cherchye et al. (2015c)). However, a crucial difference is that we allow for joint production and thereby relax the assumption of perfect weak separability.[3]

Moreover, we use this methodology to compute the efficiency of households' consumption and time use decisions in monetary terms. Towards this end, we adopt an idea put forward by Afriat (1973) and Varian (1990), i.e. the Critical Cost Efficiency index. The Critical Cost Efficiency index computes the amount by which each observed budget set must be perturbed in order for the observed decisions to be optimal. We employ the critical cost efficiency index as a measure of goodness-of-fit, indicating how well the theoretical model fits the observed data.

Second, our specific focus on the domestic production of child well-being requires us to follow a different approach than Graham and Green (1984), Kerkhofs and Kooreman (2003) and Gørtz (2011). Because these authors build on the original model (without process benefits) by Gronau (1977), they implicitly assume perfect substitutability between market goods and the corresponding domestically produced goods. We acknowledge that this assumption is reasonable for most domestically produced goods, however it seems unrealistic to assume that child well-being can be purchased in itself. For this reason, we no longer assume perfect substitutability for the domestically produced commodity in our setting. Moreover, Pollak and Wachter (1975) argue that it is difficult to measure child well-being as the output of a production technology. One could use different metrics as a proxy for child well-being, or combine several single metrics in an aggregate measure for child well-being. However, since it is hard to identify what 'defines' child well-being, and since the relevant evaluation criteria also depend on the preferences of the parents, both approaches are debatable. The advan-

[3]Note that latent separability, introduced by Blundell and Robin (2000), is another generalization of weak separability. Latent separability also weakens the assumption of mutually exclusive commodity groupings. In household production theory, latent separability corresponds to the case where consumption goods are used to produce more than one intermediate good. Each intermediate good is then produced by a part of the total amount of consumption goods. Blundell and Robin (2000) give the example of electricity, which may be used for cooking and for lighting. However, latent separability does not capture joint production. When time is an input in the household production process, it simultaneously produces a home produced good and a substitute to pure leisure. In fact, latent separability is analogous to a production setting with output-specific inputs (see Chapters 2 and 3). By contrast, the notion of process benefits corresponds to a setting with joint inputs.

tage of our approach is that neither prices nor the outputs related to child well-being must be observed. The revealed preference conditions are well-suited to deal with unknown shadow prices and domestic commodities, while it is still possible to impose restrictions on the domestic production technology. We apply our model to data that is drawn from the LISS (Longitudinal Internet Studies for the Social sciences) panel administered by CentERdata (Tilburg University, The Netherlands). Our test only requires information on the time use of both partners, their individual wages, and the allocation of (market) goods among the household members. This information is readily available in the LISS panel.

Third, we extend our nonparametric test of process benefits to a collective setting, developed by Chiappori (1988, 1992). By now, the collective model has become a standard model to analyze decisions of households on time use and consumption. In contrast to the unitary setting, which assumes that households behave as if they were a single decision maker, the collective setting explicitly recognizes that households consist of multiple decision makers, each having their own preferences. In the first part of the chapter, we develop a unitary test of the household labor supply model with home production and process benefits. Consequently, we extend our test to a collective setting. In a collective setting, we consider the well-being of the children as a public good in the household (Blundell et al. (2005) and Cherchye et al. (2012)).

Outline In Section 2, we present a labor supply model that incorporates home production and process benefits related to child care. Although our model can, in principle, deal with process benefits associated with all possible household activities (or even market work), we focus on child care. In Section 3, we derive a revealed preference test for consistency with the newly proposed model, in a unitary setting. We show that the conditions have an intuitive interpretation. Section 4 presents our data sample from the LISS panel and gives summary statistics for all relevant variables, including expenditures, individual wages and time use information. In Section 5 we identify the level of process benefits necessary to maximize the goodness-of-fit of our general model. In Section 6, we introduce process benefits in the collective model, as a final robustness check. Section 7 concludes.

4.2 Theory

In this section, we demonstrate how process benefits can be incorporated in a (unitary) labor supply model. We first present a baseline labor supply model with home production but without process benefits. Next, we formally introduce process benefits and their effect on the household's preference ordering.

Unitary labor supply with domestic production Households decide on how to allocate the available time and goods within the household. In our model, parents allocate their time among market work m^i, child care h^i and leisure l^i. For each parent i the time constraint is therefore equal to

$$h^i + l^i + m^i = T^i.$$

Each hour of market work yields a wage w^i. Hence, the amount of market work of both parents determines the household budget constraint:

$$q + X = w^1 m^1 + w^2 m^2 + y.$$

The difference between total expenditures and total labor income is the (residual) non-labor income y, which can be positive or negative. The parents in the household consume a Hicksian composite good q, with a price normalized to one. Next to their own consumption, the parents invest a part X of the household budget in the children. Following Becker (1960) and Becker (1965), we model these expenditures X and the time invested in children h^1 and h^2 as inputs in a household technology that produces child well-being $Q(h^1, h^2, X)$.

Finally, we assume that parents also value their leisure. This gives the following household level utility function

$$U(q, Q(h^1, h^2, X), l^1, l^2),$$

in which the first argument is the Hicksian composite good q, the second argument Q is child well-being, and the final arguments are leisure.

Note that we assume separability between the decisions on private consumption, private leisure and child well-being on the one hand and other household-related decisions on the other hand. We assume that some exogenous amount of time is allocated over market work, leisure and child care, and that this allocation is independent of other categories of time use. As a result, the total amount of time T^i is individual specific.

Moreover, we implicitly assume that households maximize the above objective function as if they were a single decision maker. In Section 6, we will relax this assumption by building on the collective model by Chiappori (1988, 1992) and Apps and Rees (1988).

Although the above expression nests a production function within the household's utility function, it is still a fairly typical representation of a household's preference ordering in a classical labor supply setting. We will extend this setting by allowing for process benefits.

Process benefits Specifically, we take into account that parents can 'enjoy' spending time with their children. In such case, each unit of time h^i invested in child care jointly produces indirect utility (through the utility of the child) and direct utility in terms of leisure. Indeed, we will add a specific fraction of h^i to the pure leisure l^i of parent i. We express this relationship between total leisure L^i, pure leisure l^i and child care h^i as follows:

$$L^i(l^i, h^i) = l^i + g^i(h^i),$$

where $g^i(h^i)$ gives the additional leisure-equivalent time that enters the utility function as a result of child care time h^i. Towards this end, we follow Graham and Green (1984) and Kerkhofs and Kooreman (2003) and propose the following functional form:

$$g^i(h^i) = h^i - \frac{1}{(T^i)^{\delta^i}} \frac{(h^i)^{1+\delta^i}}{1+\delta^i}. \tag{4.1}$$

The attractive feature of the functional specification g^i is that the shape of this function is fully determined by the parameter δ^i. Figure 4.1 provides insight in the evolution of $g^i(\cdot)$ conditional on δ^i. A setting without process benefits corresponds to a δ^i equal to 0: child care is not valued as leisure. By contrast, for δ^i approaching to infinity, each hour of child care is almost fully valued as leisure.

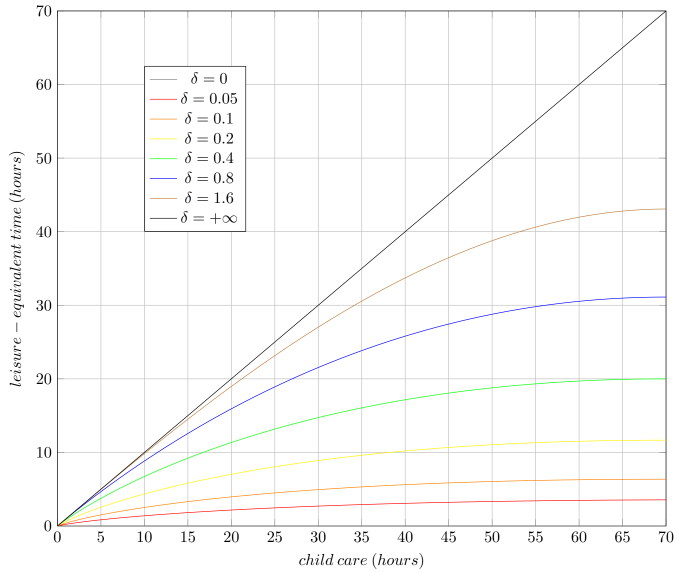


Figure 4.1: Leisure-equivalent function for different specifications of δ^i

The function g^i satisfies the following properties:

1. $0 \leq \frac{\partial g^i}{\partial h^i}(h^i) \leq 1$
2. $\frac{\partial^2 g^i}{\partial h^2}(h^i) < 0$
3. $\lim_{h^i \to 0} \frac{\partial g^i}{\partial h^i}(h^i) = 1$
4. $\lim_{h^i \to T^i} \frac{\partial g^i}{\partial h^i}(h^i) = 0.$

The intuition behind these properties is the following. For a positive value of δ^i, the first hour of child care is fully valued as an hour of leisure, whereas each additional hour of child care is valued less so. The extent to which this decreases is determined by the parameter δ^i: the parameter δ^i captures the intensity of the process benefits. We can now set up a new household objective function

$$U(q, Q, L^1, L^2) = U(q, Q(h^1, h^2, X), l^1 + g^1(h^1), l^2 + g^2(h^2)),$$

in which the leisure arguments are augmented with process benefits $g^1(h^1)$ and $g^2(h^2)$, respectively. We could also allow for multiple kinds of process benefits. For example child care h^i gives process benefits $g^i(h^i)$ and market work m^i gives process benefits $k^i(m^i)$ such that total leisure $L^i = l^i + g^i(h^i) + k^i(m^i)$. In such a framework, it is possible that market work is more enjoyable than child care. However, let us focus on the process benefits associated with child care in this chapter.

At this point, it might be worth noting that in the original work by Becker (1965), different time uses have different values. In practice, the value of a person's different time uses is not observed. A typical solution is to assume that an individual's different time uses have a uniform price, which is equal to the individual's market wage. This is also the way we handle the opportunity cost of child care and leisure. However, as Cherchye et al. (2015a) and Chiappori and Lewbel (2015) argue, this leads to a fundamental identification problem and it is impossible to disentangle preferences from technologies. This would bring us back to the traditional labor supply model, in which all non-market time can be aggregated into a single composite 'leisure'. However, by the way we include household production and process benefits into the model, we impose additional structure on the opportunity cost of child care time. As we will show in the empirical application, different specifications of process benefits do not lead to observationally equivalent models.

Unitary labor supply with domestic production and process benefits We focus on households with two parents ($i = 1, 2$) and one or more children. According to the (unitary) labor supply model, households maximize a joint objective function, subject to the household budget constraint.

Problem 4.1. *Optimization problem $U - PB$*

$$\max_{q,h^1,h^2,l^1,l^2,X} U(q, Q(h^1, h^2, X), l^1 + g^1(h^1), l^2 + g^2(h^2))$$
$$\text{s.t. } q + X + w_s^1(h^1 + l^1) + w_s^2(h^2 + l^2) \leq w_s^1 T_s^1 + w_s^2 T_s^2 + y_s$$

To optimize Problem 4.1, households decide on q, h^1, h^2, l^1, l^2 and X. Suppose we have a data set $\mathscr{S} = \{[q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2], [w_s^1, w_s^2] | s \in S\}$ in which we observe these decisions. We consider the observed decisions to be rational if there exists a utility function U and a production function Q such that the observations $(q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2)$ solve optimization problem $U - PB$. Alternatively to the notion rationality, we will also use the notion of consistency with a particular model, to make explicit which behavioral model we have in mind when discussing rationality. The data set $\mathscr{S}$ is said to be consistent with the unitary labor supply model with home production and process benefits.

Definition 4.1. *Consider a data set* $\mathscr{S} = \{[q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2], [w_s^1, w_s^2] | s \in S\}$. *We say that the data set* $\mathscr{S}$ *is consistent with the unitary labor supply model with home production and process benefits* g^i *if there exist a concave utility function* U *and a concave production function* Q *such that the observed decisions solve optimization problem* $U - PB$.

We assume that the observed choices in $\mathscr{S}$ are the results of the decision making process of households with the same preferences and the same household production function. In the empirical application, we work with a cross section of households. To test consistency of observed choices, we will divide households in the sample into household types, on the basis of observable characteristics. We then assume that households belonging to the same household type have the same preferences and household production function.

4.3 Revealed preference

In this section, we introduce necessary and sufficient conditions to test whether the decisions of a number of households $s \in S$ can be described as rational choices, conditional on some specification of the process benefits g^1 and g^2. Towards this end, we use revealed preference methodology and we impose no structure on the household's utility function U and production function Q. We simply test the joint hypothesis of rational behavior and process benefits captured by g^1 and g^2. The results will be independent of any parametric specification of preferences or the household's production technology.

This leads to a sharp test of rationality: behavior is either rational or not rational. Next, we relax our conditions to test whether household behavior is 'nearly' optimizing under the assumption of some level of process benefits. In particular, we allow households to make a small optimization error and thereby waste a small fraction of the household budget. We use an adaptation of Afriat's Critical Cost Efficiency Index, which has an intuitive monetary interpretation.

The tests that we discuss allow us to verify consistency of a model with the data. We will use these methods in Section 5 to test the rationality (efficiency) of the households' decisions given some specification of process benefits. To be more specific, we assume that the model with the largest efficiency index fits the data best. This allows us to examine which specification of process benefits is the most appropriate. Furthermore, we also discuss the discriminatory power of the tests in Section 5.

4.3.1 Setting the stage: individual labor supply

Before we derive testable revealed preference conditions for Definition 4.1, we present the more simple revealed preference conditions associated with an individual labor supply model (without home production and without process benefits). Consider for instance a setting where an individual chooses to allocate her time over market work m and leisure l. Utility is derived from leisure l and from a good q (with price p) purchased with the labor income $w \cdot m$. By choosing (q_s, l_s) the individual solves

$$\max_{q,l} U(q,l) \text{ s.t. } p_s q + w_s l \leq w_s T_s.$$

Suppose there exist $s \in S$ observations of price-quantity pairs $(p_s, w_s; q_s, l_s)$. These $|S|$ observations can correspond to the individual's labor supply in $|S|$ different time periods (under the assumption of stable preferences) or to the labor supply decisions of $|S|$ different individuals with homogeneous preferences. To investigate whether there exists a utility function $U(\cdot,\cdot)$ that rationalizes $\mathscr{S} = \{[q_s, l_s], [p_s, w_s] | s \in S\}$[4], one can use Afriat's Theorem.[5]

Proposition 4.1. *For a given data set $\mathscr{S} = \{[q_s, l_s], [p_s, w_s] | s \in S\}$, the following statements are equivalent:*

[4] We assume that the set $\mathscr{S}$ is a finite set. For infinite data sets, we refer to Reny (2015) for equivalent conditions. In particular, there not necessarily exists a utility function which is concave and continuous, if the data set satisfies GARP. Concavity and continuity are in the infinite case replaced by weaker assumptions.

[5] We refer to seminal contributions by Samuelson (1938), Afriat (1967), Diewert (1973) and Varian (1982).

1. *There exists a locally non-satiated utility function that rationalizes the data set $\mathscr{S}$.*

2. *There exists a continuous, concave and monotone utility function that rationalizes the data $\mathscr{S}$.*

3. *There exist $\lambda_s \in \mathbb{R}_{++}$ and $u_s \in \mathbb{R}_+$ such that $\forall s, v \in S$:*

$$u_s - u_v \leq \lambda_v(p_v(q_s - q_v) + w_v(l_s - l_v)).$$

4. *The data set $\mathscr{S}$ satisfies GARP.*

Conveniently, this equivalence theorem provides us with two practical tests for the existence of a (continuous, concave and monotone) utility function that rationalizes the data. First, rationality requires that there exist strictly positive variables λ_s and positive utility numbers u_s such that the so called Afriat inequalities hold. The inequalities combine the first-order conditions (where U_{q_v} and U_{l_v} are the marginal utilities from private consumption and leisure, respectively),

$$U_{q_v} = \lambda_v p_v \text{ and } U_{l_v} = \lambda_v w_v$$

with concavity of the utility function.

Alternatively, rationality requires that $\mathscr{S} = \{[q_s, l_s], [p_s, w_s] | s \in S\}$ satisfies the Generalized Axiom of Revealed Preference (GARP). Essentially, the GARP constructs combinatorial restrictions by which rationality can be tested. First, GARP requires that if $p_s q_s + w_s l_s \geq p_s q_v + w_s l_v$ then bundle (q_s, l_s) is directly revealed preferred over bundle (q_v, l_v) (or formally, $(q_s, l_s)\ R_0\ (q_v, l_v)$). Indeed, (q_v, l_v) was affordable at prices s but yet not chosen. Second, GARP imposes that the (direct) revealed preference relations are transitive. If $(q_s, l_s)\ R_0\ (q_v, l_v)\ R_0\ (q_z, l_z)$ then $(q_s, l_s)\ R\ (q_z, l_z)$ because bundle (q_s, l_s) is (indirectly) revealed preferred over bundle (q_z, l_z). Finally, GARP states that when $(q_s, l_s)\ R\ (q_z, l_z)$, it should not be the case that $p_z q_s + w_z l_s < p_z q_z + w_z l_z$ because when bundle s is preferred to bundle z there is no reason to purchase bundle z at prices (p_z, w_z).

4.3.2 Revealed preference analysis of labor supply with home production and process benefits

In this paragraph, we derive the revealed preference conditions for a (unitary) labor supply model with home production and process benefits. These conditions allow to test whether households behave as rational agents given that the household members experience certain levels of process benefits. At this point, it is worth to emphasize that the test is conditional on a particular specification of the process benefits. We have chosen to do so for practical reasons. When the function g^i is specified, one can compute the values $g^i(h_s^i)$ and $\frac{\partial g^i}{\partial h}(h_s^i)$. Otherwise, the testable conditions would always remain nonlinear and hard to check in practice. In the empirical application, we will check a variety of parameter values for the function g^i and select the values that give the best empirical fit of the model to the data.

Before presenting the testable conditions, we first compute the first-order conditions associated with Problem 4.1. A more extensive discussion is available in Appendix 4.B. The first-order conditions associated with q and l simply impose that

$$U_{q_s} = \lambda_s \text{ and } U_{L_s^i} = U_{l_s^i} = \lambda_s w_s^i,$$

with λ_s the Lagrange multiplier. Let us then focus on the first-order conditions associated with home production. We obtain that

$$U_{Q_s} Q_{h_s^i} + U_{L_s^i} \frac{\partial g^i}{\partial h_s^i} = \lambda_s w_s^i \text{ and } U_{Q_s} Q_{X_s} = \lambda_s.$$

In these conditions, U_{q_s}, U_{Q_s}, U_{L_s} and U_{l_s} give the marginal utility with respect to q_s, Q_s, L_s and l_s. $Q_{h_s^i}$ is the marginal productivity of child care by parent i and Q_{X_s} is the marginal productivity of money spent on children. In words, the latter condition shows that the marginal willingness to pay for child well-being U_{Q_s}/λ_s is equal to the price of an additional unit of child well-being $(1/Q_{X_s})$ in equilibrium. Rewriting these equations gives

$$\frac{Q_{h_s^i}}{Q_{X_s}} = w_s^i(1 - \frac{\partial g^i}{\partial h_s^i}). \tag{4.2}$$

This is exactly the condition for an optimal (i.e. cost minimizing) allocation of inputs applied to our setting. The only difference with respect to a labor supply model without process benefits is the $(1 - \partial g^i / \partial h_s^i)$ factor. This factor decreases the relative input price of child care to the extent that the parent i enjoys child care. In the limit, when process benefits are maximal ($\partial g^i / \partial h_s^i \to 1$), it may be rational to care for children even if this is almost not productive (i.e. if $Q_{h_s^i} \to 0$).

We finally combine these restrictions with concavity of the utility function and concavity of the production function. This gives the following proposition.

Proposition 4.2. *For a given data set $\mathscr{S} = \{[q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2], [w_s^1, w_s^2] | s \in S\}$, the following statements are equivalent:*

1. *The data set $\mathscr{S}$ is consistent with the unitary labor supply model with home production and process benefits g^1 and g^2.*

2. *There exist shadow prices $\wp_s \in \mathbb{R}_{++}$, production levels $Q_s \in \mathbb{R}_+$, utility numbers $u_s \in \mathbb{R}_+$ and multipliers $\lambda_s \in \mathbb{R}_{++}$ such that $\forall s, v \in S$:*

   (a) $u_v - u_s \leq \lambda_s \left((q_v - q_s) + \wp_s (Q_v - Q_s) + w_s^1 (L_v^1 - L_s^1) + w_s^2 (L_v^2 - L_s^2) \right)$

   (b) $\wp_s (Q_v - Q_s) \leq w_s^1 (1 - \frac{\partial g^1}{\partial h_s^1})(h_v^1 - h_s^1) + w_s^2 (1 - \frac{\partial g^2}{\partial h_s^2})(h_v^2 - h_s^2) + (X_v - X_s)$

   with $L_s^i = L^i(l_s^i, h_s^i) = l_s^i + g^i(h_s^i)$.

3. *There exist shadow prices $\wp_s \in \mathbb{R}_{++}$ and production levels $Q_s \in \mathbb{R}_+$, such that $\forall s, v \in S$:*

   (a) *the data set $\{(1, \wp_s, w_s^1, w_s^2; q_s, Q_s, L_s^1, L_s^2)\}$ satisfies GARP*

   (b) $\wp_s (Q_v - Q_s) \leq w_s^1 (1 - \frac{\partial g^1}{\partial h_s^1})(h_v^1 - h_s^1) + w_s^2 (1 - \frac{\partial g^2}{\partial h_s^2})(h_v^2 - h_s^2) + (X_v - X_s)$

   with $L_s^i = L^i(l_s^i, h_s^i) = l_s^i + g^i(h_s^i)$.

This gives a necessary and sufficient test of consistency with Definition 4.1. The set of conditions (2a) is formally similar to the standard Afriat inequalities. It results from combining concavity (of the utility function) and the first-order conditions of the optimization problem. The set of conditions (2b) stems from concavity of the production function. We refer to Appendix 4.A for a proof.

In contrast to the standard Afriat inequalities in Proposition 4.1, the second set of conditions is not easily verified. The difficulty is that the prices $\wp_s$ and corresponding quantities Q_s of child well-being are both unobserved. This problem is inherent to the 'child well-being' good under consideration because it is nonmarketable (its 'price' is unobserved) and hard to measure (its 'quantity' is unobserved). We have already discussed in the introduction that the parametric approach to home production, as initiated by Gronau (1977), is less useful for studying nonmarketable goods. An attractive feature of our nonparametric approach is that prices or outputs of the domestically produced good need not be observed. However, this leads to a set of nonlinear conditions due to the factors $\wp_s(Q_v - Q_s)$.

To obtain testable conditions, we use Afriat's theorem to translate condition (2a) in terms of the Generalized Axiom of Revealed Preference (see supra) and obtain condition (3a). We indicate in Appendix 4.B how conditions (3a) and (3b) can be tested by solving a mixed integer linear programming problem. The revealed preference relations are then captured by binary variables R_{sv} which have a value equal to one if and only if $u_s \geq u_v$.

4.3.3 Empirical issues

To end this section, we discuss three empirical issues which are relevant for the interpretation of our results in Section 4.5: the goodness-of-fit of our model, shadow price restrictions and the non-nestedness of different specifications of process benefits.

Goodness-of-fit The test in Proposition 4.2 is a 'sharp' test of rationality: a data set is either rationalizable or not. However, such a sharp test is not always informative. Following Afriat (1973) and Varian (1990) we allow 'nearly' rational decision makers to make small optimization errors. We therefore relax the conditions in Proposition 4.2 such that a household is allowed to waste a part of the household budget. Before relaxing Proposition 4.2, note that the inclusion of process benefits into the model has consequences for the opportunity cost of time. Intuitively, in a situation in which process benefits are present, a choice of child care time h_s^i and pure leisure l_s^i, at price w_s^i and process benefits g^i, could be interpreted as if the parent chose the combination of child care time h_s^i and equivalent leisure L_s^i, with the opportunity cost of time respectively equal to $(1 - \frac{\partial g^i}{\partial h_s^i})w_s^i$ and w_s^i. To allow households to waste a part of the budget, we will relax the conditions in Proposition 4.2 by replacing the chosen bundle $(q_s, X_s, h_s^1, h_s^2, L_s^1, L_s^2)$ by the adjusted quantity bundle $(eq_s, eX_s, eh_s^1, eh_s^2, eL_s^1, eL_s^2)$. Note that we keep prices, and therefore also the opportunity cost $w_s^i(1 - \frac{\partial g^i}{\partial h_s^i})$ of child care time for the production of child well-being, fixed to the original level. A household is said to be rationalizable at efficiency level e if the household wastes not more than a fraction $(1 - e)$ of the budget by making irrational choices.

Definition 4.2. *The dataset $S = \{[q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2], [w_s^1, w_s^2] | s \in S\}$ is consistent with the unitary labour supply model with home production and process benefits g^1 and g^2, at efficiency level e, if there exist shadow prices $\wp_s \in \mathbb{R}_{++}$, production levels $Q_{es} \in \mathbb{R}_+$ and $Q_s \in \mathbb{R}_+$ ($Q_{es} \leq Q_s$), utility numbers $u_s \in \mathbb{R}_+$ and multipliers $\lambda_s \in \mathbb{R}_{++}$ such that the following conditions hold for all s, v in S:*

$$u_v - u_s \leq \lambda_s \left((q_v - eq_s) + \wp_s(Q_v - Q_{es}) + w_s^1(L_v^1 - eL_s^1) + w_s^2(L_v^2 - eL_s^2)\right)$$

$$\wp_s(Q_v - Q_{es}) \leq w_s^1(1 - \tfrac{\partial g^1}{\partial h_s^1})(h_v^1 - eh_s^1) + w_s^2(1 - \tfrac{\partial g^2}{\partial h_s^2})(h_v^2 - eh_s^2) + (X_v - eX_s)$$

with $L_s^i = l_s^i + g^i(h_s^i)$

A small technical remark is in order. We introduce an additional variable Q_{es}, which corresponds to $Q(eX_s, eh_s^1, eh_s^2)$. We impose that $Q_{es} \leq Q_s$, since we assume that Q is an increasing function. Although Definition 4.2 is rather technical, the efficiency level e has an intuitive interpretation in terms of money waste. For an efficiency level of $e = 1$, the conditions are equivalent to the original consistency conditions in Proposition 4.2. If the original consistency conditions are satisfied, the behavior is fully rational and no money is wasted due to suboptimal decision making. If a dataset can be rationalized at efficiency level e, with $0 \leq e \leq 1$, an implication of Definition 4.2 is that if $u_v \geq u_s$, then

$$\begin{aligned} & q_v + w_s^1(1 - \tfrac{\partial g^1}{\partial h_s^1})h_v^1 + w_s^2(1 - \tfrac{\partial g^2}{\partial h_s^2})h_v^2 + w_s^1 L_v^1 + w_s^2 L_v^2 + X_v \\ \geq \quad & e\left(q_s + w_s^1(1 - \tfrac{\partial g^1}{\partial h_s^1})h_s^1 + w_s^2(1 - \tfrac{\partial g^2}{\partial h_s^2})h_s^2 + w_s^1 L_s^1 + w_s^2 L_s^2 + X_s\right). \end{aligned}$$

Suppose that a household prefers the bundle $(q_v, q_v^2, X_v, h_v^1, h_v^2, L_v^1, L_v^2)$ over the bundle $(q_s, X_s, h_s^1, h_s^2, L_s^1, L_s^2)$ $(u_v \geq u_s)$. If bundle v would have been affordable at time s, then choosing bundle s would not be a fully rational choice, since bundle v results into a higher utility and was even cheaper. Consequently, money would have been wasted by choosing bundle s. In line with Definition 4.2, the households are allowed to waste a small part of the budget by making irrational choices. In particular, if the dataset is rationalizable at efficiency level e, then the condition is that bundle v is not more than $(1 - e) \cdot 100$ % cheaper than bundle s.

We define the critical cost efficiency e_c as the largest efficiency e at which the household behavior is rationalizable. To determine the critical cost efficiency e_c in practice, we first test whether the conditions are satisfied for $e = 1$. If not, we test smaller and smaller values of e, until we find e_c for which the conditions are satisfied. The critical cost efficiency e_c provides us with an attractive measure of how well the model fits the data. If the critical cost efficiency is low, the households seem to have wasted a large part of the budget. This indicates that the observed decisions are not in line with the theoretical model we have in mind. Stated differently, the theoretical model does not fit the data very well in that case. In the remainder of the chapter, we use the critical cost efficiency as our goodness-of-fit measure. This goodness-of-fit measure quantifies if the model under study is a reasonable model to study the observed behavior.

Shadow price restrictions We have already argued that the price of child well-being is unobserved. This domestically produced good is nonmarketable, which makes it impossible to attach a price to one 'unit' of child well-being. We use shadow price $\wp_s$ to measure the willingness to pay for an additional unit of child well-being. This $\wp_s$ is only restricted by the conditions in Proposition 4.2 because utility functions remain unspecified. However, we argue that 'extreme' variation in this shadow price is unrealistic. First of all, we assume that households care for the well-being of the children and therefore $\wp_s \neq 0$. If households would not care at all, this would mean that the home production channel is trivial and that it provides no natural benchmark for our extension with process benefits. Second, in the application we will select households which are similar in terms of observed characteristics. For these reasons, we restrict the variation in $\wp_s$. The maximum value of $\wp_{\max}$ in a pool of households can not exceed twice the size of the minimum value $\wp_{\min}$ in the same pool. This restriction is similar in spirit to the revealed preference restrictions on the variation in the intrahousehold distribution of bargaining power across (non-unitary) households, see e.g. Bruyneel et al. (2012).

Non-nestedness We want to emphasize that models with varying levels of process benefits are not nested. Raising the level of process benefits does not necessarily leads to conditions that are easier to satisfy. The reason is that process benefits impact on our conditions in two (opposite) ways. In particular, if δ increases, then both $g^i(h^i)$ and $\frac{\partial g^i}{\partial h^i}$ increase. On the one hand, increasing $g^i(h^i)$ increases the total amount of equivalent leisure L^i in the first condition. On the other hand, increasing $\frac{\partial g^i}{\partial h^i}$ reduces the opportunity cost of child care $(1 - \frac{\partial g^i}{\partial h^i})w^i$ in the second condition. Consequently, if a larger δ tightens the first condition, the second condition is strengthened and vice versa.

4.4 Data

We apply our nonparametric methodology to data from the (Dutch) Longitudinal Internet Studies for the Social sciences (LISS). The LISS panel is a representative sample of Dutch individuals who participate in monthly Internet surveys.[6] This data source contains information on household characteristics, wages, private expenditures, expenditures on children, and the allocation of time over leisure, home production (child care) and market work. The relevant data is organized in different modules. We use data from the Background Variables, Family and Household, Work and Schooling and Time Use and Consumption. Note that the latter module was specifically added to investigate consumption and time use decisions within the household. The information in this module is collected in three waves: 2009, 2010 and 2012. Cherchye et al. (2012) constructed their sample from the same modules. Because wages are not directly available, we re-construct the wages by dividing the monthly net labor income by the (average) number of hours worked.

[6] The panel is based on a true probability sample of households drawn from the population register. Households that could not otherwise participate are provided with a computer and Internet connection. A longitudinal survey is fielded in the panel every year, covering a large variety of domains including work, education, income, housing, time use, political views, values and personality.

Let us first explain how we selected the sample. We started by selecting all households with 1, 2 or 3 children.[7] Consequently, we dropped all households with missing household head and/or wedded or unwedded partner and all households with members other than parents or children. Moreover, observations with missing values for wages, consumption decisions or time use choices were withdrawn. We end up with a total of 470 households.

Our sample, collected from the LISS panel, consists of couples with children, for which both partners participate in the labor market. We therefore have a sample with very specific characteristics. In 2013, the average number of children in the Netherlands is equal to 1.68. This is slightly less than the average of two children in our sample, however this may not be a surprise, since we only selected couples with at least one child. Furthermore, our sample only consists of couples. Statistics Netherlands reports that about 20% of the households with children living at home is a one-parent family. The strongest selection criteria might be the fact that both parents need to participate in the labor market. We need both parents to participate, since we need the market wage as an indication for the opportunity cost of time. Statistics Netherlands (cbs) reports figures of women with children who have paid work in the Netherlands. Cloïn (2013) reports that the percentage of married (or cohabiting) mothers with underage children, that has paid work has been increasing over time, from 60% in 2002 to 72% in 2013. Furthermore, the percentage of mothers that work, depends on the number of children in the household. In 2013, 82% of the mothers with 1 child work, 78% of the mothers with 2 children and 61% of the women with three children. When interpreting our results on process benefits, we need to keep in mind that our results are conditional on labor market participation. Although we restricted our analysis to couples for which both partners work, the situation in which the mother works part time and the father works full time is the most commonly observed in the Netherlands. In Section 4.5, we take a closer look at the number of hours that mothers and fathers on average work.

Table 4.1 reports summary statistics of the (nondurable) expenditures, the time use and the socioeconomic background of the households in the sample. Private expenditures consist of expenditures on eating at home, food and drinks outside the house, cigarettes and

[7] There are households with 4 or 5 children present in the LISS data. However, we did not include those in our final sample. After cleaning the data, there were too little observations to do a sound analysis of process benefits for those household types.

	Husband Mean (Std.)	Wife Mean (Std.)	Household Mean (Std.)
Expenditures (EUR per week)			
Private Expenditures	79,7 (56,0)	89,4 (70,4)	
Expenditures on Children			142,6 (124,6)
Time use (hours per week)			
Market labor (incl commuting time)	46,9 (11,7)	30,3 (11,7)	
Child Care	11,8 (9,2)	17,8 (13,8)	
Leisure	25,7 (13,7)	24,0 (13,4)	
Socioeconomic variables			
Wage rate (EUR per hour)	12,6 (4,0)	11,6 (3,8)	
Age	42,6 (6,6)	40,4 (6,6)	
Mean age children			9,5 (5,9)
Age variation children			1,3 (1,2)
Number of children			2 (0,7)

Table 4.1: Summary statistics of sample of 470 households, with 1, 2 or 3 children.

other tobacco products, clothing, personal care products and services, medical care and health costs not covered by insurance, leisure time expenditures, (further) schooling, donations and gifts, and other personal expenses. The expenditures on children consist of all the expenditures discussed above, but now on behalf of the children. We also add expenditures on children's daycare to the child expenditures. Finally, we normalize all expenditures to a weekly basis.

Parents allocate time to market work, home production (child care) and leisure. First of all, individual leisure is directly available from the LISS panel. Second, market work is the sum of actual labor time and time necessary for traveling from and to work. Finally, time invested in children includes bathing, dressing, playing, reading stories, going with the child to the doctor, taking the child to school or hobbies. On average, the fathers in the sample take care of their children 12 hours a week, while the mothers spend 18 hours a week taking care of their children. Note that besides the time that parents care for the children, hired daycare is also included in the analysis through the channel of the child expenditures. By contrast, the time that grand parent's take care of the children is not included.

To perform the analysis, we divided the 470 households into 32 household types, on

	#	l^1	l^2	h^1	h^2	m^1	m^2	w^1	w^2	c^1	c^2	k
All	470	25,7	24,0	11,8	17,8	46,9	30,3	12,6	11,6	79,7	89,4	142,6
1 child	109	24,8	24,7	12,8	19,7	46,7	30,4	12,3	11,6	80,7	98,3	100,0
2 children	266	26,4	24,5	11,7	16,9	46,7	30,2	12,7	11,4	80,2	87,7	141,9
3 children	95	24,7	22,0	10,8	18,2	47,9	30,3	12,7	12,2	77,1	84,3	193,4
Age $[0, 5[$	120	23,0	19,9	18,1	29,5	46,5	29,4	12,2	11,5	82,6	84,6	164,4
Age $[5, 9[$	106	27,0	24,7	12,9	20,2	46,9	29,1	12,3	11,9	78,0	89,0	137,3
Age $[9, 13[$	104	26,3	24,6	10,5	15,5	46,7	29,9	13,2	11,6	75,6	86,3	113,2
Age $[13, \infty[$	140	26,4	26,8	6,6	8,1	47,8	32,2	12,8	11,4	85,4	94,7	158,8
No high ed	212	25,0	24,3	10,4	16,2	48,1	28,9	11,2	10,6	75,5	80,7	124,9
Higher ed	258	26,2	23,9	12,8	19,0	46,2	31,4	13,8	12,3	85,2	95,7	161,6

Table 4.2: Time use and consumption by household type (1=husband, 2=wife)

the basis of observable characteristics. Within each household type, it is assumed that the preferences as well as the home production technology are homogeneous. First, we focus on characteristics related to the children. After all, the number of children in a household and the children's age strongly affect the appropriate household technology. Our sample consists of households with 1, 2 or 3 children. There are 109 households with 1 child, 266 households with 2 children and 95 households with 3 children. Furthermore, we distinguish between 4 age categories: households with children aged less than 5 years (on average), households with children aged between 5 and 9 years, households with children aged between 9 and 13 years old and households with children aged more than 13 years. These boundaries on the age categories are convenient for us, since they lead to a subdivision of the full sample into four age categories of more or less equal size. The younger the children, the more time the parents spend caring for the children. For example, in households with children aged less than 5 years, father and mother care on average 18 and 30 hours a week, respectively. This in contrast to households with older children (average age more than 13). In these households, the parents spend respectively 7 and 8 hours a week caring for the children.

Since there are many households with 2 children in our sample, we divided these households into even more specific household types, on the basis of the variation in the age of the children. In particular, we distinguish between households with little variation in the age of the children and households with a lot of variation in age.[8] Our final criterion to subdivide the sample was the education of the parents. In particular, we distinguish between households with at least one adult having higher education (higher vocational education or

[8] Our threshold is the median of the standard deviation of the age of the children, for households with 2 children.

university) and households without higher education. This subdivision divides our sample more or less in half. Our motivation to subdivide the sample on the basis of education is recent work of Guryan et al. (2008), which studies the relationship between parents' earnings potential and their time spent with children, in a cross-section from the American Time Use Surveys. These authors' results indicated that highly educated parents, in particular, are more likely to spend time with their children. This is surprising, because the opportunity cost of child care by a highly educated individual is considerably higher. We refer to Guryan et al. (2008) for possible explanations of this observation. Based on this evidence, we constructed household types with and without higher education. In this way, we avoid the need to assume that households with and without higher education have the same preferences for child well-being. In our sample, 38% of the husbands is highly educated and 39% of the women is highly educated. To conclude, we divided the households into 32 household types, on the basis of the number of children, the age of the children and the education level of the parents. Table 4.2 reports summary statistics of the time use and consumption of the households types. The first column of the table reports the number of households in the sample that satisfies each criterion. A pool of households having the same household type consists on average of 15 households, with a minimum of 6 households and a maximum of 28 households. We refer to Table 4.8 in Appendix 4.D for a detailed overview of the household types.

At this point it is worth to note that we have done several robustness checks to test whether the goodness-of-fit results are robust to changes in the division of the sample into household types. For example, we constructed household types on the basis of the age of the youngest child, instead of the mean age of the children. Next, we also experimented with the number of age categories. Furthermore, we tested whether taking up additional characteristics, such as the gender of the children and the age of the parents had a large effect on the goodness-of-fit results and the gains in terms of goodness-of-fit from including process benefits into the model. The average goodness-of-fit was each time very similar to the results we will present in the following section.

4.5 Results

As a first step, we compute the goodness-of-fit of the model with and without process benefits for each household type. We assume that couples within a household type have the same preferences U and the same household production function Q. In Section 4.3.2, we presented a test for process benefits, under the assumption that the fathers and the mothers belonging to the same household type perceive the same intensity of process benefits. However, we can perfectly relax this assumption and allow the parameter δ_s^i to be individual-specific. The intensity of the process benefits may then not only differ from father to mother, but may differ from household to household. From now on, each individual may have a different perception of child care as leisure. To select the optimal value of the process benefits, we select parameter values that maximize the goodness-of-fit of the model to the data. In a next step, we study the distribution of the estimated process benefits. Finally, we also discuss the discriminatory power of the test with and without process benefits.

Process benefits: specification and testing To select optimal values for the process benefits, we consider a set of values $\{0, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6\}$. These values describe a heterogeneous range of process benefits. The seven chosen scenarios are clearly distinct and informative. A setting without process benefits corresponds to a $\delta^i = 0$, while setting $\delta^i = 1.6$ corresponds a to setting in which child care time is almost fully valued as leisure. Note that there is no big difference between setting $\delta^i = 1.6$ and setting δ^i even higher, because the curve associated with $\delta^i = 1.6$ approximately lies on the 45-degree line for the most values of h^i in our sample.

Per pool of households we select for each individual i in each household s a parameter value δ_s^i from the set $\{0, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6\}$. We perform a random search for the optimal parameter values. In particular, we consider 10000 combinations of δ_s^i drawn at random from this set and compute for each combination of parameter values the goodness-of-fit. The goodness-of-fit is assessed by using the critical efficiency score e_c, which quantifies deviations from optimal decision-making. We then select the combination of parameter values δ_s^1 and δ_s^2 which provides the best empirical fit e^*. At this point, it might be worth to note that there usually exist multiple combinations of parameter values that rationalize the observed behavior. If we find multiple combinations of parameter values having an equally good empirical fit e^*, we select the combination which gives the 'minimal' level of process benefits.[9] In this way we can interpret our process benefits as small but necessary deviations from the traditional labor supply model without process benefits. Note that a search with 10000 random draws does not test every possible combination of parameter values. However, since we allow δ_s^i to be individual specific, a grid search on the parameter values would suffer from the curse of dimensionality. Still, the improvements in terms of goodness-of-fit are already substantial by considering 10000 combinations of parameter values. Note that considering a larger number of combinations, or considering a finer grid, could lead to even more gains from process benefits.

In Table 4.3, we present some summary statistics of the goodness-of-fit of the model with and without process benefits for each household type. For the model with process benefits, we report the best empirical fit e^* in the search for parameter values for the process benefits. The last row gives the gains from process benefits, which are computed as the difference between the goodness-of-fit e^*, when process benefits are taken into account and the goodness-of-fit e^0 with respect to the model without process benefits.

		min	mean	median	max
Goodness-of-fit with process benefits	e^*	0,75	0,93	0,96	1,00
Goodness-of-fit without process benefits	e^0	0,65	0,87	0,90	1,00
Gains from Process benefits	$e^* - e^0$	0	0,06	0,04	0,19

Table 4.3: Summary statistics of goodness-of-fit for each of the 32 household types

A critical cost efficiency score e indicates that a share $(1 - e)$ of the household's resources has been wasted due to sub-optimal decision making. Under the model without process benefits, it seems as if the observed households waste on average 13% of their resources due to sub-optimal decision making. In particular, if the households would have chosen a different combination of consumption and time use, they could have chosen a combination which was equally good in terms of utility, and which was up to 13% cheaper

[9] In particular, if there are multiple combinations of parameter values having an equally good empirical fit e^*, we compute for each individual $\frac{g^i(h_s^i)}{h_s^i}$, which is the proportion of child care time perceived as leisure. We then select the parameter values that imply the minimal average proportion within a pool, necessary to achieve efficiency e^*.

than the currently chosen combination. Under the model that does allows process benefits, the observed households waste on average 7% of their resources. Stated differently, the goodness-of-fit of the model with process benefits is on average 6% higher than the model without. We conclude that including process benefits substantially improves the goodness-of-fit of the model. We refer to Table 4.9 in the appendix for an overview of the results per household type. There is quite some variation in the gains from process benefits. On the one hand, there are 5 household types that do not benefit from including process benefits into the model. On the other hand, the maximal gains from including process benefits is 19% in terms of goodness-of-fit.

Table 4.4 reports summary statistics of the optimal values for the process benefits δ_s^i, i.e. the parameter values corresponding to the best empirical fit.

	min	mean	median	max
Process benefits father δ^1	0	0,31	0,10	1,60
Process benefits mother δ^2	0	0,31	0,10	1,60
Proportion father	0	0,35	0,24	1,00
Proportion mother	0	0,32	0,21	1,00

Table 4.4: Summary statistics of estimated values of process benefits and proportion of child care time perceived as leisure, for father and mother in each of the 470 households.

The average process benefit parameters δ^1 and δ^2 across all households equal 0.31, both for fathers and mothers.[10] Given our optimal parameter values and the functional specification of process benefits in (4.1), we can recover the marginal and average leisure from child care. The marginal leisure $\frac{\partial g(h^i)}{\partial h^i}$ can be used to interpret the opportunity cost of child care time in equilibrium. Intuitively, it is less costly to invest time in children for parents that experience more process benefits.

However, we focus our discussion here on the average leisure $\frac{g(h^i)}{h^i}$, which indicates the proportion of child care time that is perceived as leisure. Table 4.4 reports some summary statistics. We find that the proportion of child care perceived as leisure varies considerably across individuals. The smallest level is 0 (i.e. no process benefits) whereas the largest level

[10]To test the robustness of our results, we repeated the random grid search of 10000 iterations several times. The average goodness-of-fit and therefore also the gains from process benefits were very robust. Each time we obtained an average goodness-of-fit equal to 0.93. Although the individually selected parameter values for the process benefits could differ, the average process benefits in the sample for mother and father differed not more than a couple of percentage points.

is 1 (i.e. 100 $\%$ of time spent with children is valued as leisure).

We present the distribution of the proportion of child care perceived as leisure across all individuals in Figure 4.2, distinguishing between fathers (full curve) and mothers (dashed curve).

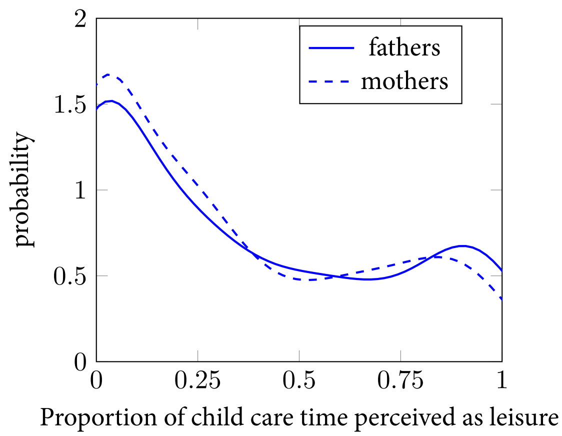


Figure 4.2: Probability density of proportion of child care perceived as leisure

The corresponding probability density curves show a peak at zero or very small process benefits. The large concentration of observations around 0 can be explained by the fact the goodness-of-fit of several household types can not be improved by including process benefits into the model. Although the average level of process benefits equals 0.31 for both mothers and fathers, the average percentage a child care time perceived as leisure is slightly larger for fathers than for mothers (35% for fathers, 32% for mothers). This difference is also visible in Figure 4.2. The logical explanation for this is the fact that fathers spend on average less hours a week caring for the children than the mothers (respectively 12 and 18 hours a week). The assumptions we made about the functional specification of the process benefits implies that the first hour of child care is (almost) fully perceived as leisure and every subsequent hour less and less. The extent to which this decreases is determined by the intensity δ^i of the process benefits. As a consequence of our functional assumption, the fathers perceive on average a (slightly) larger proportion of their child care time as leisure -even though the intensity of process benefits we estimate is on average the same. However, this difference in the proportion of child care perceived as leisure is not significant. A Kolmogorov-Smirnov test of equal distributions accepts the null hypothesis that the male and female proportion of child care perceived as leisure follow the same distribution ($p - value = 0.32$).[11]

Contractual working hours As mentioned before, we selected a sample from the LISS panel, in which both parents participate on the labor market. When interpreting our results on process benefits, we therefore need to keep in mind that our results are conditional on labor market participation. However, one could argue that the decision on whether or not to participate in the labor market, depends on the presence of children in the household. Parents might feel the need to work less hours in the labor market, in order to have enough time to care for the children. Moreover, parents who genuinely enjoy taking care of the children and doing household related chores, might be more likely to work less hours in the labor market. Furthermore, Del Boca (2002) find that the availability of child care and part time opportunities increase both the probability of working and having a child. There is no need to say that the relation between the labor market participation decision, the children in the household and the degree of process benefits is a complex issue. We refer to Browning (1992) for an interesting discussion on children and household economic behavior. In particular, Browning (1992) also discuss labor supply models with endogenous fertility. Moreover, Kalb (2009) discusses challenges for an empirical analysis of children, labor supply and child care. An interesting avenue for future research is to investigate the relation between the labor market participation decision and the degree of process benefits related to child care. In this paper we make abstraction of the labor market decision. As a preliminary exercise, we investigate the amount of process benefits, conditional on the hours of employment. Note that in Section 4.4 we reported the weekly hours of market labor, including commuting time, as reported in the time use study. Here we take a closer look at the contractual working hours. Table 4.5 reports summary statistics of the number of contractual working hours. In our sample, the number of contractual hours that fathers in our

[11] It is sometimes argued in the literature that, on average, "mothers care more for children than fathers," in the sense that an increase in the mother's power within the couple results in more expenditures made for children, see Blundell et al. (2005) and references therein. Blundell et al. (2005) propose a theoretical model to investigate the claim that mothers care more for children than fathers. Cherchye et al. (2012), who build further on this model, find no evidence that empowering mothers is more beneficial for the well-being of the children than empowering fathers. Similarly, we find no evidence that mothers enjoy caring for the children more than fathers.

sample work is on average equal to 38,8 hours per week. Furthermore, the average hours of contractual work for mothers is equal to 25,5 hours per week. Note that these numbers are relatively close to the national statistics in the Netherlands, reported by Statistics Netherlands (cbs). For example, in the year 2013, men and women in the Netherlands which are member of a couple and with under-age children, work on average respectively 40,2 and 24,4 hours per week. Conditional on the fact that both parents work, there does not seem to be a large difference in our sample in working hours between parents with more or less and younger or older children.

Table 4.5: Child care, contractual work and household chores by household type (hours per week)

	Child care		Work		Household chores	
	Father	Mother	Father	Mother	Father	Mother
All	11,8	17,8	38,8	25,2	9,3	16,9
1 child	12,8	19,7	37,5	26,3	9,6	15,3
2 children	11,7	16,9	39,2	24,8	9,1	17,0
3 children	10,8	18,2	39,4	25,1	9,5	18,7
Age $[0, 5[$	18,1	29,5	39,0	25,8	9,2	13,7
Age $[5, 9[$	12,9	20,2	38,2	24,2	8,6	16,7
Age $[9, 13[$	10,5	15,5	39,0	24,8	9,4	18,6
Age $[13, \infty[$	6,6	8,1	38,9	26,1	9,8	18,7
No high ed	10,4	16,2	39,9	23,7	8,5	17,3
Higher ed	12,8	19,0	37,9	26,5	10,0	16,6

We use the median values in the sample, which are equal to 38 hours per week for fathers and 24 hours per week for mothers, to divide the mothers and fathers in two groups. Consequently, we investigate the amount of process benefits, conditional on the hours of employment. Figure 4.3 plots the probability density of the proportion of child care that fathers perceive as leisure, for fathers working either less than 38 hours per week, or more than 38 hours per week. A Kolmogorov Smirnov test (p-value = 0.33) reports that the distribution of process benefits of those two groups of fathers working more or less hours is not significantly different. Similarly, we do not find a significant difference between the process benefits for mothers who work more than 24 hours per week, and mothers who work less than 24 hours per week (p-value = 0.21), see Figure 4.4.

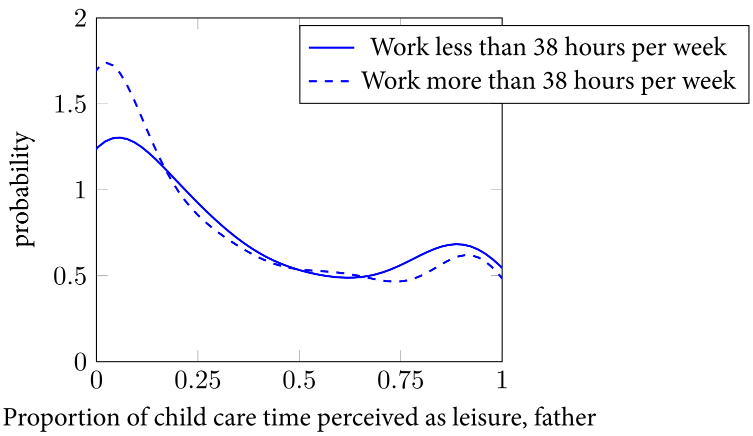


Figure 4.3: Child care time, conditional on contractual working hours, father

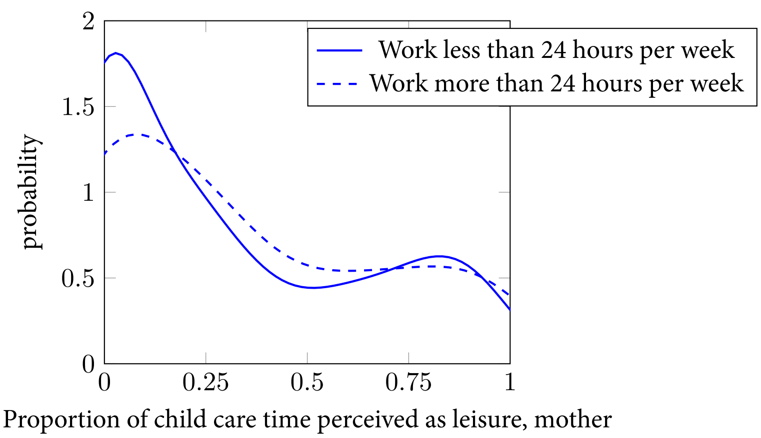


Figure 4.4: Child care time, conditional on contractual working hours, mothers

Household chores We assumed separability between the decisions on private consumption, private leisure and child well-being on the one hand and other household related decisions on the other hand. Consequently, we have assumed that the choice to invest time in child care is independent of the choice to spend time on household chores. Table 4.5 reports the number of hours per week that father and mother in our sample spend on household chores, such as cleaning, laundry, shopping, cooking,... On average, fathers and mothers spend respectively $9,3$ and $16,9$ hours per week on household chores. The difference between fathers and mothers corresponds to findings of Cloïn (2013), who report that women in the Netherlands nearly spend twice as much time to household chores than men. Table 4.5 also reports the average hours of household chores by household type. For fathers, the average per type is very stable over all types. For mothers, we observe a slight increase in household chores when the number of children and the age of the children increases. As a robustness check for our separability assumption, we test whether the level of process benefits depends on the amount of time spend on household chores. We again subdivide the fathers and mothers into two groups, depending on whether they spend more or less time on household chores than the median father of mother.

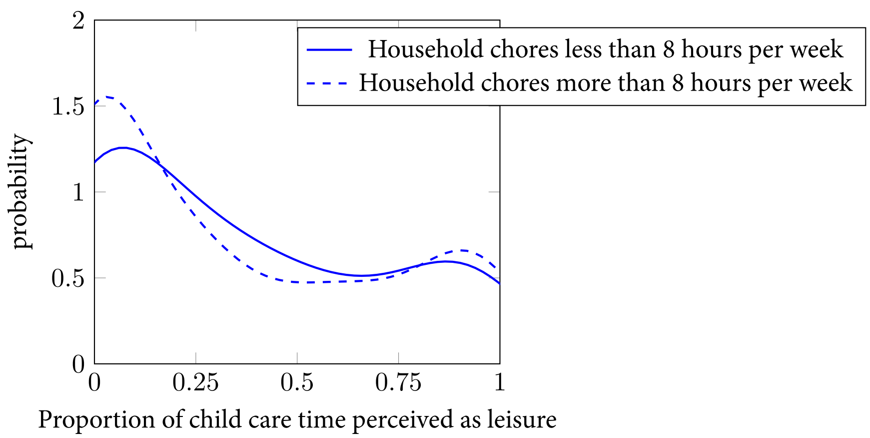


Figure 4.5: Child care time, conditional on hours of household chores, father

Figure 4.5 and Figure 4.6 show the probability density of the proportion of child care perceived as leisure, for the two groups of fathers and mothers. However, we do not find a significant difference on the basis of the amount of time spend on household chores. In particular, the p-values equal respectively 0.21 and 0.87 for fathers and mothers. An interesting avenue for future research is to include other types of household activities into the analysis of process benefits. In principle, various types of process benefits could be associated with various time uses. For example, it might not be unreasonable to argue that time spent

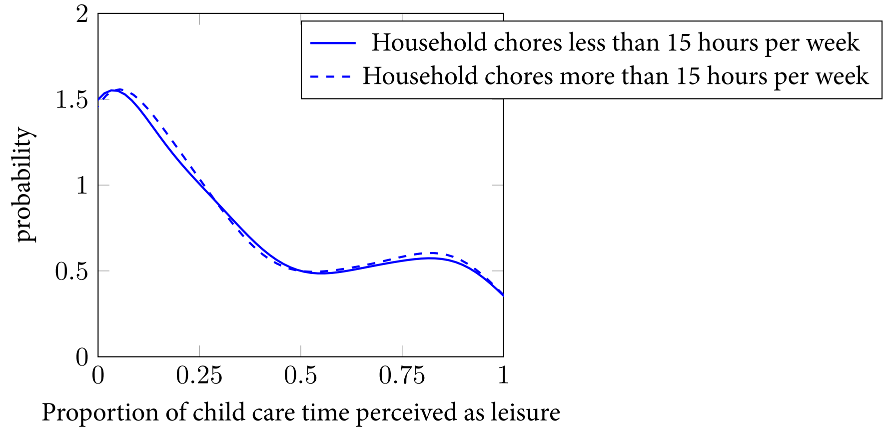


Figure 4.6: Child care time, conditional on hours of household chores, mother

cooking or cleaning can also be a direct source of utility or disutility for the household.[12]

Discriminatory power It is well-known that revealed preference tests sometimes lack 'discriminatory power', i.e. that they are unable to reject the consistency of totally random, irrational choices. This may occur when there is insufficient price variation in a data set, or the underlying model is too weak. In our setting we fully exploit interhousehold heterogeneity in wages, which provides considerable price variation. However, the question remains whether allowing for process benefits does not weaken our testable implications. In other words, we examine whether the optimal specification of process benefits - which maximizes the goodness-of-fit - is still powerful enough to discriminate between observed rational choices and simulated irrational choices. We have already argued that different specifications of our model are not necessarily nested. Allowing for higher levels of process benefits can either weaken or strengthen our testable conditions. In particular, higher levels of process benefits may relax the restrictions associated with optimal production (of child well-being) but simultaneously tighten the conditions associated with optimal con-

[12] In theory, other household produced goods could be included in the analysis, and process benefits related to various home production activities could be included. In practice, we can only handle one type of home produced good, otherwise, we create nonlinearities in the formulation of the household optimization problem, which makes the theoretical model difficult to test. This is a limitation of our nonparametric approach and the fact that we do not observe the output of the home produced good. If we would have data on the output of an additional home produced good, we would perfectly be able to include process benefits related to this home produced good.

sumption. The final outcome then depends on the trade-off between the effect of a lower (opportunity) cost of child care and the effect of a higher level of leisure-equivalent time.

To quantify the discriminatory power of our test, we simulate 1000 random data sets per household type. One such data set consists of random consumption and time use choices for all households in the pool. The consumption choices correspond to random draws (of budget shares) from a uniform distribution on the unit interval. We then divide these simulated shares by the corresponding prices and multiply the result with the observed total expenditures per household. Furthermore, we simulate random time use choices by drawing time allocations from a uniform distribution on the unit interval. This is similar to Bronars (1987)'s power index, which is based on Becker (1962)'s notion of irrational behavior.

		Min	Mean	Median	Max
No	Goodness-of-fit observed data	0,65	0,87	0,90	1,00
process	Goodness-of-fit random data	0,67	0,85	0,87	0,97
benefits	Difference observed vs random	-0,14	0,02	0,02	0,23
Best	Goodness-of-fit observed data	0,75	0,93	0,96	1,00
process	Goodness-of-fit random data	0,41	0,67	0,64	0,94
benefits	Difference observed vs random	0,02	0,26	0,31	0,41

Table 4.6: Differences in goodness-of-fit of observed versus random choices

In Table 4.6 we compare the goodness-of-fit of actual choices and the (average) goodness-of-fit of random choices. Table 4.6 reports some summary statistics, we refer to Table 4.10 in the appendix for a detailed overview of the goodness-of-fit for every household type. The higher the (average) goodness-of-fit of the random choices, the lower the discriminatory power of our test.

Ideally, the goodness-of-fit of observed choices is much higher than those of randomly simulated choices. Let us first investigate the performance of the labor supply model *without* process benefits. On average, the observed choices are slightly more efficient than random choices (average efficiency of 0.87 versus 0.85). For some household types, the model without process benefits even gives a better fit for the random data pools than for the observed data pool. We then turn to the performance of the labor supply model with best process benefits. This model conditions on the level of process benefits that gives the highest efficiency to the observed data set. It turns out that this alternative model performs much better than the labor supply model without process benefits. The efficiency of observed choices is always higher, with differences ranging from 0.02 to 0.41. This indicates that the discriminatory power of the model with process benefits is much higher than the model without process benefits. These results confirm that our extension is reasonable.

What do we learn from all this? The results in this section are important from a theoretical, methodological and empirical perspective. Theoretically, the results motivate the extension of standard labor supply models (with home production) with process benefits. We have found that mothers and fathers perceive part of their time with children as leisure. Methodologically, we have shown how to test and identify home production models - with process benefits - in which the outputs are not only unobserved but also unmarketable. Our model is able to distinguish between observed choices and randomly simulated choices. This confirms that our method has 'empirical bite', even when preferences and the home production technology remain unspecified. Empirically, we found considerable variation in the intensity of the process benefits from individual to individual. However, we did not find indications that the intensity of process benefits differs between fathers and mothers.

4.6 Extension: collective labor supply

As a robustness check, we extend our nonparametric test of process benefits to a collective setting. The collective model was developed by Chiappori (1988, 1992) and is by now standard in the literature. In contrast to the unitary setting, which assumes that households behave as if they were a single decision maker, the collective setting explicitly recognizes that households consist of multiple decision makers, each having their own preferences. After all, we can reasonably expect the preferences of fathers to differ from the preferences of mothers. As soon as the bargaining power between fathers and mothers varies, this might lead to a violation of the assumption that households maximize a single, homogeneous utility function. The collective model lets the relative bargaining power vary across households within a pool.

Note that we do consider the parents as two individual decision makers, but still not the children. There is some discussion in the literature on the role of children in the decision making process of the household (e.g. Dauphin et al. (2011)). This discussion is particularly relevant for older children (age 16 and older), however it is less debatable that younger children are bystanders in the household. To model the presence of (younger) children in the household, children are often treated as a public good in the household (e.g. Blundell et al. (2005) and Cherchye et al. (2012)). This is the approach we will follow.

Similar to the unitary model, we can introduce process benefits in the collective model. To implement process benefits in the collective model, we build on contributions of Chiappori (1997) and Blundell et al. (2005), who introduced household production in the collective model. For compactness, we refer to Appendix 4.C for more details on the conditions to test consistency with the collective model with household production of child well-being and with process benefits. We must note that the test in Appendix 4.C is a necessary test for collective rationality.[13] However, the collective test is not always more easily satisfied than the unitary test. For example, the intra-household allocation q_1 and q_2 of private goods plays a role in the collective model. This in contrast to the unitary test, which only takes the sum $q = q_1 + q_2$ of the private goods into account. The data sample retrieved from the Longitudinal Internet Studies for the Social sciences, contains information on the private consumption of husband and wife.

As a robustness check, we investigate whether the extension of the home production model with process benefits still adds value in the collective framework. Stated differently, does including process benefits into the collective model still leads to gains in terms of goodness-of-fit? We therefore apply the test in Proposition 4.3 in Appendix 4.C to our data sample from the Longitudinal Internet Studies for the Social sciences. Towards this end we again run a search for the parameter values δ^i in the grid $\{0, 0.05, 0.1, 0.2, 0.4, 0.8, 1.6\}$ and select the combination of process benefits δ^i which gives the best 'goodness-of-fit'. The results are presented in Table 4.7.

The average goodness-of-fit within the sample equals 0.98, which is 5% higher than the

[13] In theory, it is perfectly possible to construct a test for the collective model with home production and process benefits which is both necessary and sufficient. However, this test can not be reformulated in terms of linear programming conditions. We refer to Appendix 4.C for more explanation. We therefore limit the analysis to a necessary test of the model, which is slightly weaker than the necessary and sufficient test. The necessary test can be easily implemented using standard linear programming techniques.

		min	mean	median	max
Goodness-of-fit with process benefits	e_c^*	0,87	0,98	1,00	1,00
Goodness-of-fit without process benefits	e_c^0	0,76	0,92	0,93	1,00
Gains from Process benefits	$e_c^* - e_c^0$	0	0,06	0,04	0,20

Table 4.7: Summary statistics of goodness-of-fit for each of the 32 household types, for the collective labor supply model with home production

unitary test. Remarkably, the gains in terms of goodness-of-fit from including process benefits in the model remain 6%, the same as for the unitary case. For some household types, the gains from including process benefits into the model reach up to 20%. Our results indicate that it is important to deal with process benefits, also in collective labor supply models. We refer to Table 4.9 and Table 4.11 in Appendix 4.D for an overview of the goodness-of-fit results and discriminatory power for every household type.

The discriminatory power of the collective model with and without process benefits is comparable to that of the unitary model. In particular, the goodness-of-fit of the collective model without process benefits is on average 1% higher for the observed data than for the random simulated data. By contrast, the goodness-of-fit of the collective model with process benefits is on average 20 % higher for the observed data. These findings strongly confirm our results for the unitary model.

4.7 Conclusion

We presented tests of the labor supply model with home production of child well-being and process benefits. By allowing for process benefits, we take into account that parents may also enjoy spending time with children, and perceive some fraction of child care time as individual leisure. In this way, child care time 'jointly' produces well-being for the household's children - modeled as domestic production - and leisure time for the caring parent - modeled as a substitute for pure leisure. Standard labor supply models ignore this possibility.

Our contribution to the literature is threefold. First, we identify the minimal level of process benefits necessary to rationalize the choices in a sample of Dutch households, from the Longitudinal Internet Studies for the Social sciences (LISS). This minimal level is independent of parametric assumptions on the household members' preferences and the household's home production technology. As a result our method is very robust to mis-specification of these underlying, and generally unknown, elements. Second, our method can deal with home production even if the corresponding output is not only unobserved but also unmarketable. This makes that our methodology is particularly useful for studying the well-being of children. After all, child well-being is hard to observe and its price is hard to measure. Metrics for child well-being are typically debatable. At a more general level, our nonparametric framework with unobserved child outputs could be used as a benchmark to test the fit of different metrics of child well-being. The conditions of our nonparametric model - which only imposes monotonicity and concavity of the household's technology function - may even be strengthened by replacing the unobserved output levels Q with some metric for child well-being. The corresponding effect on the goodness-of-fit of the model then indicates whether the metric is reasonable, or not. To implement process benefits in this model of home production, we generalize the revealed preference conditions in Cherchye et al. (2015c) to take 'joint production' into account. The conditions in Cherchye et al. (2015c) allow to test consistency with the maximization of a (single) utility function which is also weakly separable in some arguments. Our extension is necessary because process benefits make that utility is no longer fully weakly separable in child well-being. After all, time spent with children enters both the domestic production technology and the utility function(s) - as a substitute for leisure. Finally, we propose a necessary test for process benefits in the collective model. This allows us to deal with very general forms of interhousehold variation in the bargaining power of households.

Our results indicate that including process benefits significantly improves the 'goodness-of-fit' of labor supply models with home production. This is consistent with the findings by Juster (1985), Hallberg and Klevmarken (2003) and Krueger et al. (2009), who argue that the home production of child well-being is one of the more enjoyable activities in the household. Although we find considerable variation in the intensity of process benefits, we find no evidence that fathers or mothers intrinsically perceive child care as more or less enjoyable. On a more general level, the substantial interhousehold heterogeneity in process benefits also underlines the importance of 'joint' production as a possible driver of labor supply and child care decisions.

We see different avenues for further research. First, we exploit the cross-sectional dimension of the LISS data for two reasons. On the one hand, this gives us considerable wage variation, which improves the empirical bite of our conditions. On the other hand, our empirical analysis requires data on both consumption and time use, which severely limits the number of households in the sample. However, panel data would allow researchers to pool consumption and time use choices per household. This eliminates the issues associated with interhousehold variation in preferences. Second, we have restricted attention to a static model of consumption and time use. The domestically produced child well-being depends on current expenditures and current time use choices. It would be interesting to investigate how our framework extends to a more general intertemporal setting. Third, it is possible to impose additional structure on the elements underlying the households' decision problems. Structure on the domestic technology that 'produces' child well-being may shed light on the productivity of parental time inputs as well as on the substitutability between mothers' and fathers' time invested in children. A particularly convenient parametric specification for the household's production function is proposed in Lise and Yamada (2014). The authors' production function combines an aggregate Cobb Douglas production function, with expenditures and (aggregate) child care time as inputs, and a Constant Elasticity of Substitution production function, which produces an intermediary output (child care time) from the combination of mothers' and fathers' time invested in child care. This production function allows to investigate whether mothers' time (resp. fathers' time) spent with children is substitutable by fathers' time (resp. mothers' time) invested in children. Finally, we believe that this study motivates further research on the magnitude of process benefits and the relationship with observed family characteristics.

4.A Proofs

Proof of Proposition 4.2. We prove that the second set of conditions in Proposition 4.2 is a necessary and sufficient condition for consistency with the labor supply model with home production and process benefits.

1. **Necessity.** We have that each observed $(q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2)$ solves the problem

$$\max_{q,X,h^1,h^2,l^1,l^2} U(q, Q(h^1, h^2, X), l^1 + g^1(h^1), l^2 + g^2(h^2))$$

$$s.t.$$

$$q + X + w_s^1(h^1 + l^1) + w_s^2(h^2 + l^2) \leq w_s^1 T_s^1 + w_s^2 T_s^2.$$

Given concavity, the functions U and Q are superdifferentiable. Consequently, the objective function is superdifferentiable. Define λ_s as the Lagrange multiplier associated with the constraint. An optimal interior solution for the above maximization problem must satisfy the following first order conditions:

$$U_{q_s} = \lambda_s \tag{4.3}$$

$$U_{Q_s} Q_{h_s^i} + U_{L_s^i} \frac{\partial g^i}{\partial h_s^i} = \lambda_s w_s^i \tag{4.4}$$

$$U_{Q_s} Q_{X_s} = \lambda_s \tag{4.5}$$

$$U_{L_s^i} = \lambda_s w_s^i \tag{4.6}$$

for f_{x_s} a superderivative of the concave function f defined for x and evaluated at x_s. Here is f equal to U or Q and x_s equal to q_s, Q_s, L_s^i, h_s^i or X_s. Next, define $\wp_s = \frac{1}{\lambda_s} U_{Q_s}$.

This gives

$$\wp_s = \frac{1}{Q_{X_s}}, \tag{4.7}$$

$$\wp_s = \frac{w_s^i(1 - \frac{\partial g^i}{\partial h_s^i})}{Q_{h_s^i}}. \tag{4.8}$$

Next, denote $U(q_s, Q_s, L_s^1, L_s^2) = u_s$ for each s in S. Then the concavity of the function U in q, Q and L^i implies

$$u_v - u_s \leq U_{q_s}(q_v - q_s) + U_{Q_s}(Q_v - Q_s) + \sum_i U_{L^i_s}(L^i_v - L^i_s)$$

for v, s in S. Combining the first order conditions with the above inequality, we obtain

$$u_v - u_s \leq \lambda_s \left((q_v - q_s) + \wp_s(Q_v - Q_s) + \sum_i w^i_s(L^i_v - L^i_s) \right).$$

Concavity of the function Q in h^1, h^2 and X implies

$$Q_v - Q_s \leq Q_{h^1_s}(h^1_v - h^1_s) + Q_{h^2_s}(h^2_v - h^2_s) + Q_{X_s}(X_v - X_s).$$

Applying equations (4.7) and (4.8), results into the condition

$$\wp_s(Q_v - Q_s) \leq w^1_s(1 - \frac{\partial g^1}{\partial h^1_s})(h^1_v - h^1_s) + w^2_s(1 - \frac{\partial g^2}{\partial h^2_s})(h^2_v - h^2_s) + (X_v - X_s),$$

which proves the revealed preference conditions.

2. **Sufficiency.** Suppose that the revealed preference conditions hold.

   For any $(\hat{q}, \hat{X}, \hat{h}^1, \hat{h}^2, \hat{l}^1, \hat{l}^2, \hat{Q})$ such that

$$\hat{q} + \sum_{i=1,2} \left(w^i_s(\hat{l}^i + \hat{h}^i) \right) + \hat{X} \leq q_s + \sum_{i=1,2} \left(w^i_s(l^i_s + h^i_s) \right) + X_s,$$

   we can define

$$Q(\hat{h}^1, \hat{h}^2, \hat{X}) = \min_{v \in S} \left[Q_v + \frac{1}{\wp_v} \left(z^1_v(\hat{h}^1 - h^1_v) + z^2_v(\hat{h}^2 - h^2_v) + P_v(\hat{X} - X_v) \right) \right],$$

   with

$$z^i_v = w^i_v(1 - \frac{\partial g^i}{\partial h^i_v}).$$

Assume that $Q(h_s^1, h_s^2, X_s)$ reaches its minimum at observation n:

$$Q(h_s^1, h_s^2, X_s) = Q_n + \frac{1}{\wp_n}\left(z_n^1(h_s^1 - h_n^1) + z_n^2(h_s^1 - h_n^2) + P_n(X_s - X_n)\right).$$

Moreover, the production inequalities imply that

$$Q_n + \frac{1}{\wp_n}\left(z_n^1(h_s^1 - h_n^1) + z_n^2(h_s^1 - h_n^2) + P_n(X_s - X_n)\right) \geq Q_s.$$

Hence, $Q(h_s^1, h_s^2, X_s) \geq Q_s$ which shows that Q_s gives a lower bound on $Q(h_s^1, h_s^2, X_s)$.

Likewise,

$$U(\hat{q}, \hat{Q}, \hat{L}^1, \hat{L}^2) = \min_{v \in S}\left[u_v + \lambda_v\left((\hat{q} - q_v) + \wp_v(\hat{Q} - Q_v) + \sum_i w_v^i(\hat{L}^i - L_v^i)\right)\right].$$

Assume that $U(q_s, Q_s, L_s^1, L_s^2)$ reaches its minimum at observation m:

$$U(q_s, Q_s, L_s^1, L_s^2) = u_m + \lambda_m\left((q_s - q_m) + \wp_m(Q_s - Q_m) + \sum_i w_m^i(L_s^i - L_m^i)\right). \tag{4.9}$$

Moreover, the Afriat-like inequalities imply that

$$u_m + \lambda_m\left((q_s - q_m) + \wp_m(Q_s - Q_m) + \sum_i w_m^i(L_s^i - L_m^i)\right) \geq u_s. \tag{4.10}$$

Combining equations (4.9) and (4.10) gives

$$\begin{aligned} & U(q_s, Q_s, L_s^1, L_s^2) \\ = \;& u_m + \lambda_m\left((q_s - q_m) + \wp_m(Q_s - Q_m) + \sum_i w_m^i(L_s^i - L_m^i)\right) \\ \geq \;& u_s, \end{aligned}$$

which shows that u_s gives a lower bound on $U(q_s, Q_s, L_s^1, L_s^2)$.

By definition,

$$U(\hat{q},\hat{Q},\hat{L}^1,\hat{L}^2) \leq u_s + \lambda_s \left((\hat{q}-q_s) + \wp_s(\hat{Q}-Q_s) + \sum_i w_s^i(\hat{L}^i - L_s^i) \right) \text{ and}$$

$$\hat{Q} \leq Q(\hat{h}^1,\hat{h}^2,\hat{X}) \leq Q_s + \frac{1}{\wp_s}\Big(z_s^1(\hat{h}^1 - h_s^1) + z_s^2(\hat{h}^2 - h_s^2) + (\hat{X} - X_s) \Big).$$

Moreover, concavity of g^i implies that

$$g^i(\hat{h}^i) - g^i(h_s^i) \leq \frac{\partial g^i}{\partial h_s^i}(\hat{h}^i - h_s^i),$$

such that

$$\begin{aligned} & U(\hat{q}, Q(\hat{h}^1,\hat{h}^2,\hat{X}), \hat{l}^1 + g^1(\hat{h}^1), \hat{l}^2 + g^2(\hat{h}^2)) \\ \leq \quad & u_s + \lambda_s \Big((\hat{q}-q_s) + \big(z_s^1(\hat{h}^1 - h_s^1) + z_s^2(\hat{h}^2 - h_s^2) + (\hat{X} - X_s) \big) \\ + \sum_i \quad & w_s^i \big(\hat{l}^i - l_s^i + \frac{\partial g^i}{\partial h_s^i}(\hat{h}^i - h_s^i) \big) \Big). \end{aligned}$$

Rearranging terms and using the definition of z_s^i, we obtain

$$\begin{aligned} & U(\hat{q}, Q(\hat{h}^1,\hat{h}^2,\hat{X}), \hat{l}^1 + g^1(\hat{h}^1), \hat{l}^2 + g^2(\hat{h}^2)) \\ \leq \quad & u_s + \lambda_s \Big((\hat{q}-q_s) + \sum_{i=1,2} \Big(w_s^i(\hat{l}^i - l_s^i) + w_s^i(\hat{h}^i - h_s^i) \Big) + (\hat{X} - X_s) \Big). \end{aligned}$$

Due to the budget constraint, we have

$$\begin{aligned} & U(\hat{q}, Q(\hat{h}^1,\hat{h}^2,\hat{X}), \hat{l}^1 + g^1(\hat{h}^1), \hat{l}^2 + g^2(\hat{h}^2)) \\ \leq \quad & u_s \\ \leq \quad & U(q_s, Q(h_s^1,h_s^2,X_s), l_s^1 + g^1(h_s^1), l_s^2 + g^2(h_s^2)), \end{aligned}$$

which proves that the bundle $(q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2, Q_s)$ maximizes the utility function $U(q_s, Q(h_s^1,h_s^2,X_s), l_s^1 + g^1(h_s^1), l_s^2 + g^2(h_s^2))$ subject to the budget constraint $\hat{q} + \sum_{i=1,2}\Big(w_s^i(\hat{l}^i + \hat{h}^i)\Big) + \hat{X} \leq (q_s + \sum_{i=1,2}\big(w_s^i(l_s^i + h_s^i)\big) + X_s)$. Consequently, we constructed a production function Q and utility function U that satisfy the requested properties, such that the observed decisions solve optimization problem U-PB. Therefore, the data set S is consistent with the (unitary) labor supply model with home production and process benefits g^i.

$\square$

4.B Testable conditions

As discussed in Section 4.3.2, the Generalized Axiom of Revealed preference can be used to obtain testable conditions. To verify GARP, the conditions in Proposition 4.2 can be reformulated in terms of mixed integer programming (MIP). The binary variables R_{sv} then capture the revealed preference relations R, as defined in Section 4.3.1. Here $R_{sv} = 1$ should be interpreted as $(q_s, Q_s, L_s^1, L_s^2)\ R\ (q_v, Q_v, L_v^1, L_v^2)$.

Then the data set $\mathscr{S} = \{[q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2], [w_s^1, w_s^2] | s \in S\}$ is consistent with the unitary labor supply model with home production and process benefits g^1 and g^2 if and only if the following MIP problem is feasible:

Problem 4.2. *There exist variables β_s, production levels Q_s and binary variables R_{sv} such that the following conditions hold for all s, t, v in S:*

$$M_s R_{sv} \geq \beta_s \left(q_s + w_s^1 L_s^1 + w_s^2 L_s^2\right) + Q_s - \beta_s \left(q_v + w_s^1 L_v^1 + w_s^2 L_v^2\right) - Q_v \tag{4.11}$$

$$R_{st} + R_{tv} \leq 1 + R_{sv} \tag{4.12}$$

$$M_v(1 - R_{sv}) \geq \beta_v \left(q_v + w_v^1 L_v^1 + w_v^2 L_v^2\right) + Q_v - \beta_v \left(q_s + w_v^1 L_s^1 + w_v^2 L_s^2\right) - Q_s \tag{4.13}$$

$$Q_v - Q_s \leq \beta_s (w_s^1 (1 - \frac{\partial g^1}{\partial h_s^1})(h_v^1 - h_s^1) + w_s^2 (1 - \frac{\partial g^2}{\partial h_s^2})(h_v^2 - h_s^2) + (X_v - X_s)) \tag{4.14}$$

with $L_s^i = L^i(l_s^i, h_s^i) = l_s^i + g^i(h_s^i)$. Here we let M_v be a fixed and large number.

It is sufficient that M_v is larger than the total budget of household v. First, constraint (4.14) corresponds to condition 3b in Proposition 4.2, for which we defined $\beta_s = \dfrac{1}{\wp_s}$, to obtain linear conditions. Further, constraints (4.11), (4.12) and (4.13) correspond to the

GARP conditions. The GARP conditions can be interpreted as follows. Constraint (4.11) states that if $q_s + w_s^1 L_s^1 + w_s^2 L_s^2 + \wp_s Q_s \geq q_v + w_s^1 L_v^1 + w_s^2 L_v^2 + \wp_s Q_v$, then $R_{sv} = 1$, which means that $(q_s, Q_s, L_s^1, L_s^2)\ R\ (q_v, Q_v, L_v^1, L_v^2)$. Next, constraint (4.12) imposes transitivity of the revealed preference relations: if $R_{st} = 1$ and $R_{tv} = 1$, then $R_{sv} = 1$. Finally, condition (4.13) requires that $q_v + w_v^1 L_v^1 + w_v^2 L_v^2 + \wp_v Q_v \leq q_s + w_v^1 L_s^1 + w_v^2 L_s^2 + \wp_v Q_s$ if $R_{sv} = 1$.

4.C Collective model

The collective model differs from the unitary model in that households optimize a weighted sum of individual utilities. Importantly, the weights in the objective function can vary across observations in the same pool, because they are potentially dependent on the income at the time of marriage, the relative wages of father and mother, etc. According to the collective labor supply model, households maximize a weighted sum of fathers' utility $U^1(q^1, Q, L^1)$ and mothers' utility $U^2(q^2, Q, L^2)$, subject to the household budget constraint. The well-being Q of the children is considered to be a public good within the household.

Problem 4.3. *Optimization problem $C - PB$*

$$\max_{q^i, h^i, l^i, X} \mu_s^1 U^1(q^1, Q(h^1, h^2, X), l^1 + g^1(h^1)) + \mu_s^2 U^2(q^2, Q(h^1, h^2, X), l^2 + g^2(h^2))$$
$$\text{s.t. } q^1 + q^2 + X + w_s^1(h^1 + l^1) + w_s^2(h^2 + l^2) \leq w_s^1 T_s^1 + w_s^2 T_s^2 + y_s$$

The bargaining weights μ_s^i represent the relative bargaining power of member i in observation/household s. Also note that the decision variable q is replaced with q^1 and q^2: the intra-household allocation of private goods plays a role in the collective model. We assume that the intra-household allocation of private goods q_s^1 and q_s^2 is observed.

Definition 4.3. *Consider a data set $\mathscr{S} = \{[q_s^1, q_s^2, X_s, h_s^1, h_s^2, l_s^1, l_s^2], [w_s^1, w_s^2] | s \in S\}$. We say that the data set $\mathscr{S}$ is consistent with the collective model with household production of child well-being and with process benefits g^i if there exist concave individual utility functions U^1 and U^2, a concave production function Q and Pareto weights μ^1 and μ^2 such that the observed decisions solve optimization problem $C - PB$.*

We propose a necessary test for consistency with the collective labor supply model with home production and process benefits. Similar to the corresponding test of consistency with the unitary model, this test can be implemented using standard linear programming techniques. However, in a unitary labor supply model, there was only one unobserved shadow price associated with child well-being: $\wp_s = U_{Q_s}/\lambda_s$. In the current setting, we have two possibly distinct utility functions and hence two marginal willingness' to pay for child well-being: one for fathers $\wp_s^1 = \mu_s^1 U_{Q_s}^1/\lambda_s$ and one for mothers $\wp_s^2 = \mu_s^2 U_{Q_s}^2/\lambda_s$. From the first order conditions, it follows that $\wp_s^1 + \wp_s^2 = 1/Q_{X_s}$, or equivalently, $Q_{X_s} = \dfrac{1}{\wp_s^1 + \wp_s^2}$. Given that we need to implement the variables $\wp_s^1$ and $\wp_s^2$ separately (to reconstruct individual utility functions) and jointly (expression $\dfrac{1}{\wp_s^1 + \wp_s^2}$ in the production function), it is not possible to formulate a programming problem which provides a necessary and sufficient test, with linear conditions. Fortunately, it is possible to formulate a necessary test in terms of linear programming conditions. A necessary condition for consistency with the collective model, is that the following first-order conditions hold:

$$\wp_s^1 Q_{h_s^i} + \wp_s^2 Q_{h_s^i} + \frac{\mu_s^i U_{L_s^i}}{\lambda_s}\frac{\partial g^i}{\partial h_s^i} = w_s^i\,;$$
$$\wp_s^1 + \wp_s^2 = 1/Q_{X_s}. \tag{4.15}$$

Combining this with concavity of the utility functions and concavity of the production function, we obtain the following conditions.

$$u_v^1 - u_s^1 \leq \eta_s^1\left((q_v^1 - q_s^1) + \wp_s^1(Q_v - Q_s) + w_s^1(L_v^1 - L_s^1)\right) \tag{4.16}$$
$$u_v^2 - u_s^2 \leq \eta_s^2\left((q_v^2 - q_s^2) + \wp_s^2(Q_v - Q_s) + w_s^2(L_v^2 - L_s^2)\right) \tag{4.17}$$

$$(\wp_s^1+\wp_s^2)(Q_v-Q_s) \leq w_s^1(1-\frac{\partial g^1}{\partial h_s^1})(h_v^1-h_s^1)+w_s^2(1-\frac{\partial g^2}{\partial h_s^2})(h_v^2-h_s^2)+(X_v-X_s) \tag{4.18}$$

In a final step, we divide Condition (4.18) by $(\wp_s^1 + \wp_s^2)$, and substitute the resulting equation in $(Q_v - Q_s)$ in Conditions (4.16) and (4.17). This gives

$$\begin{aligned} u_v^i - u_s^i &\leq \eta_s^i\Big((q_v^i - q_s^i) + \tfrac{\wp_s^i}{\wp_s^1+\wp_s^2}\Big(w_s^i(1 - \tfrac{\partial g^i}{\partial h_s^i})(h_v^i - h_s^i) + w_s^j(1 - \tfrac{\partial g^j}{\partial h_s^j})(h_v^j - h_s^j) \\ &+ (X_v - X_s)\Big) + w_s^i(L_v^i - L_s^i)\Big) \end{aligned}$$

with $P_s^i = \dfrac{\wp_s^i}{\wp_s^1 + \wp_s^2}$. Intuitively, P_s^i gives the share of expenditures on children that household member i is willing to finance. By definition of P_s^i we have that $P_s^1 + P_s^2 = 1$. We obtain the following proposition:

Proposition 4.3. *Consider a data set $\mathscr{S} = \{[q_s, X_s, h_s^1, h_s^2, l_s^1, l_s^2], [w_s^1, w_s^2]|s \in S\}$. If the data set $\mathscr{S}$ is consistent with the collective labor supply model with home production and process benefits g^1 and g^2 then there exist personalized prices $P_s^i \in \mathbb{R}_+$, production levels $Q_s \in \mathbb{R}_+$, utility numbers $u_s^i \in \mathbb{R}_+$ and multipliers $\lambda_s \in \mathbb{R}_{++}$ and $\eta_s^i \in \mathbb{R}_{++}$ with $i \in \{1, 2\}$ such that the following conditions hold for all s, v in S:*

1. $P_s^1 + P_s^2 = 1$
2. $u_v^1 - u_s^1 \leq \eta_s^1\Big((q_v^1 - q_s^1) + P_s^1\Big(w_s^1(1 - \frac{\partial g^1}{\partial h_s^1})(h_v^1 - h_s^1) + w_s^2(1 - \frac{\partial g^2}{\partial h_s^2})(h_v^2 - h_s^2) + (X_v - X_s)\Big) + w_s^1(L_v^1 - L_s^1)\Big)$
3. $u_v^2 - u_s^2 \leq \eta_s^2\Big((q_v^2 - q_s^2) + P_s^2\Big(w_s^2(1 - \frac{\partial g^2}{\partial h_s^2})(h_v^2 - h_s^2) + w_s^1(1 - \frac{\partial g^1}{\partial h_s^1})(h_v^1 - h_s^1) + (X_v - X_s)\Big) + w_s^2(L_v^2 - L_s^2)\Big)$
4. $(Q_v - Q_s) \leq \lambda_s\Big(w_s^1(1 - \frac{\partial g^1}{\partial h_s^1})(h_v^1 - h_s^1) + w_s^2(1 - \frac{\partial g^2}{\partial h_s^2})(h_v^2 - h_s^2) + (X_v - X_s)\Big)$

Notice that parent i's time spent with children h^i is included in the Afriat inequalities corresponding to parent $j \neq i$, and vice versa. The reason is that time spent on child care generates positive intrahousehold externalities. Consequently, a parent is willing to pay for the child care of his or her partner. Two final remarks are in order. First, similar to the conditions in the unitary setting, the conditions in Proposition 4.3 can be reformulated in terms of a linear programming problem with binary variables. Second, similar in spirit to Definition 4.2 in the unitary setting, we can introduce an efficiency level e_c, which indicates the goodness-of-fit of the observed behavior with Proposition 4.3.

4.D Results

Household Type	Pool Size	Education Parents	Number of Children	Mean Age Children	Std Age Children
H_1	13	No Higher	1	2,3	0
H_2	8	No Higher	1	6,9	0
H_3	10	No Higher	1	10,5	0
H_4	20	No Higher	1	18,9	0
H_5	9	No Higher	2	2,1	0,8
H_6	14	No Higher	2	3,2	2,0
H_7	13	No Higher	2	6,5	1,0
H_8	20	No Higher	2	6,8	2,1
H_9	8	No Higher	2	10,8	0,9
H_{10}	21	No Higher	2	10,9	1,8
H_{11}	25	No Higher	2	16,5	0,9
H_{12}	20	No Higher	2	16,1	2,8
H_{13}	7	No Higher	3	3,4	1,7
H_{14}	10	No Higher	3	6,2	2,2
H_{15}	8	No Higher	3	11,2	3,2
H_{16}	6	No Higher	3	16,7	3,7
H_{17}	28	Higher	1	1,7	0
H_{18}	13	Higher	1	6,1	0
H_{19}	6	Higher	1	10,5	0
H_{20}	11	Higher	1	18,0	0
H_{21}	28	Higher	2	2,8	0,8
H_{22}	12	Higher	2	3,5	1,6
H_{23}	17	Higher	2	6,5	0,9
H_{24}	11	Higher	2	6,1	1,9
H_{25}	15	Higher	2	10,5	1,0
H_{26}	17	Higher	2	10,5	2,0
H_{27}	15	Higher	2	16,0	0,7
H_{28}	21	Higher	2	16,3	1,9
H_{29}	9	Higher	3	3,2	2,2
H_{30}	14	Higher	3	6,4	2,1
H_{31}	19	Higher	3	10,8	2,4
H_{32}	22	Higher	3	15,8	1,9

Table 4.8: Division of sample in household types

Type	Results unitary model				Results collective model			
	Goodness of fit	Gains from process benefits	Average process benefits		Goodness of fit	Gains from process benefits	Average process benefits	
	e_u^*	$e_u^* - e_u^0$	δ_u^1	δ_u^2	e_c^*	$e_c^* - e_c^0$	δ_c^1	δ_c^2
H_1	0,98	0,01	0,27	0,28	1	0	0	0
H_2	0,99	0,1	0,40	0,48	1	0,07	0,13	0,07
H_3	0,96	0,1	0,22	0,21	1	0,06	0,17	0,10
H_4	0,75	0,06	0,29	0,41	0,96	0,12	0,46	0,54
H_5	0,98	0,14	0,72	0,51	1	0,19	0,85	0,22
H_6	0,91	0,13	0,54	0,25	0,98	0,15	0,59	0,33
H_7	0,97	0,03	0,30	0,24	1	0	0	0
H_8	0,88	0,06	0,44	0,30	0,96	0,07	0,52	0,20
H_9	1	0	0	0	1	0	0,00	0,00
H_{10}	0,88	0	0	0	0,96	0,04	0,37	0,20
H_{11}	0,93	0	0	0	0,98	0,03	0,38	0,694
H_{12}	0,86	0,15	0,48	0,50	0,95	0,17	0,12	0,65
H_{13}	1	0,13	0,66	0,36	1	0,02	0,05	0,10
H_{14}	0,99	0,04	0,45	0,66	1	0,02	0,09	0,21
H_{15}	1	0,06	0,13	0,23	1	0	0	0
H_{16}	1	0,02	0,04	0,31	1	0	0	0
H_{17}	0,82	0,03	0,44	0,30	0,94	0,01	0,31	0,25
H_{18}	0,96	0,02	0,27	0,25	1	0,01	0,25	0,31
H_{19}	1	0,01	0,11	0,04	1	0	0	0
H_{20}	0,96	0,06	0,60	0,44	1	0,04	0,32	0,11
H_{21}	0,85	0	0	0	0,87	0,01	0,31	0,46
H_{22}	0,97	0,02	0,46	0,38	1	0,12	0,24	0,39
H_{23}	0,94	0	0	0	0,99	0	0,00	0,00
H_{24}	0,95	0,05	0,45	0,88	1	0,07	0,29	0,14
H_{25}	0,91	0,07	0,49	0,40	0,99	0,06	0,36	0,19
H_{26}	0,89	0,03	0,53	0,30	0,98	0,06	0,38	0,57
H_{27}	0,89	0,14	0,17	0,49	0,99	0,2	0,39	0,36
H_{28}	0,84	0,19	0,58	0,45	0,9	0,11	0,30	0,33
H_{29}	0,97	0,03	0,69	0,17	1	0,04	0,31	0,22
H_{30}	0,98	0,01	0,33	0,36	1	0	0	0
H_{31}	0,93	0,01	0,08	0,33	0,97	0,08	0,31	0,35
H_{32}	0,83	0,18	0,36	0,64	0,93	0,17	0,22	0,57

Table 4.9: Results for the unitary and collective model with process benefits, per household type.

Household type	No process benefits			Best process benefits		
	e^0	e_r^0	$e^0 - e_r^0$	e^*	e_r^*	$e^* - e_r^*$
H_1	0,97	0,91	0,06	0,98	0,70	0,28
H_2	0,89	0,93	-0,04	0,99	0,66	0,33
H_3	0,86	0,88	-0,02	0,96	0,74	0,22
H_4	0,69	0,83	-0,14	0,75	0,54	0,21
H_5	0,84	0,80	0,04	0,98	0,62	0,36
H_6	0,78	0,89	-0,11	0,91	0,59	0,32
H_7	0,94	0,91	0,03	0,97	0,71	0,26
H_8	0,82	0,86	-0,04	0,88	0,56	0,32
H_9	1	0,94	0,06	1	0,94	0,06
H_{10}	0,88	0,86	0,02	0,88	0,86	0,02
H_{11}	0,93	0,89	0,04	0,93	0,89	0,04
H_{12}	0,71	0,74	-0,03	0,86	0,47	0,39
H_{13}	0,87	0,93	-0,06	1	0,76	0,24
H_{14}	0,95	0,91	0,04	0,99	0,66	0,33
H_{15}	0,94	0,84	0,10	1	0,79	0,21
H_{16}	0,98	0,97	0,01	1	0,92	0,08
H_{17}	0,79	0,67	0,12	0,82	0,41	0,41
H_{18}	0,94	0,88	0,06	0,96	0,66	0,30
H_{19}	0,99	0,92	0,07	1	0,88	0,12
H_{20}	0,9	0,89	0,01	0,96	0,61	0,35
H_{21}	0,85	0,74	0,11	0,85	0,74	0,11
H_{22}	0,95	0,88	0,07	0,97	0,62	0,35
H_{23}	0,94	0,83	0,11	0,94	0,83	0,11
H_{24}	0,9	0,90	0,00	0,95	0,60	0,35
H_{25}	0,84	0,84	0,00	0,91	0,55	0,36
H_{26}	0,86	0,87	-0,01	0,89	0,53	0,36
H_{27}	0,75	0,76	-0,01	0,89	0,58	0,31
H_{28}	0,65	0,78	-0,13	0,84	0,49	0,35
H_{29}	0,94	0,91	0,03	0,97	0,76	0,21
H_{30}	0,97	0,84	0,13	0,98	0,63	0,35
H_{31}	0,92	0,68	0,24	0,93	0,59	0,34
H_{32}	0,65	0,73	-0,08	0,83	0,48	0,35

Table 4.10: Goodness-of-fit of observed and random data, for the unitary model

Household type	No process benefits e^0	e^0_r	$e^0 - e^0_r$	Best process benefits e^*	e^*_r	$e^* - e^*_r$
H_1	1	0,95	0,05	1	0,95	0,05
H_2	0,93	0,97	-0,04	1	0,91	0,09
H_3	0,94	0,94	0,00	1	0,87	0,13
H_4	0,84	0,90	-0,06	0,96	0,63	0,33
H_5	0,81	0,88	-0,07	1	0,79	0,21
H_6	0,83	0,94	-0,11	0,98	0,68	0,30
H_7	1	0,96	0,04	1	0,96	0,04
H_8	0,89	0,90	-0,01	0,96	0,65	0,31
H_9	1	0,98	0,02	1	0,98	0,02
H_{10}	0,92	0,91	0,01	0,96	0,69	0,27
H_{11}	0,95	0,94	0,01	0,98	0,58	0,40
H_{12}	0,78	0,79	-0,01	0,95	0,72	0,23
H_{13}	0,98	0,97	0,01	1	0,96	0,04
H_{14}	0,98	0,96	0,02	1	0,88	0,12
H_{15}	1	0,91	0,09	1	0,91	0,09
H_{16}	1	0,99	0,01	1	0,99	0,01
H_{17}	0,93	0,75	0,18	0,94	0,57	0,37
H_{18}	0,99	0,94	0,05	1	0,79	0,21
H_{19}	1	0,97	0,03	1	0,97	0,03
H_{20}	0,96	0,94	0,02	1	0,79	0,21
H_{21}	0,86	0,80	0,06	0,87	0,57	0,30
H_{22}	0,88	0,94	-0,06	1	0,78	0,22
H_{23}	0,99	0,89	0,10	0,99	0,89	0,10
H_{24}	0,93	0,95	-0,02	1	0,83	0,17
H_{25}	0,93	0,90	0,03	0,99	0,73	0,26
H_{26}	0,92	0,92	0,00	0,98	0,65	0,33
H_{27}	0,79	0,83	-0,04	0,99	0,62	0,37
H_{28}	0,79	0,86	-0,07	0,9	0,66	0,24
H_{29}	0,96	0,96	0,00	1	0,86	0,14
H_{30}	1	0,92	0,08	1	0,92	0,08
H_{31}	0,89	0,79	0,10	0,97	0,59	0,38
H_{32}	0,76	0,80	-0,04	0,93	0,64	0,29

Table 4.11: Goodness-of-fit of observed and random data, for the collective model.

Part IV

General conclusion

The analysis of optimizing behavior - also called efficient or rational behavior - has received considerable attention over the past decades. The main objective of my dissertation was to extend and apply tools that are developed to test consistency of observed behavior with theoretically optimizing behavior. I considered optimizing behavior in two different settings, namely a production and a consumption setting. Although these two fields of study seem very different at first sight, the methods and models we discuss are both rooted in a structural model of choice behavior. The main difference between the two fields is perhaps the observability of the outputs of the production process. As a consequence, the methods to test efficient production behavior and rational consumption behavior differ in practice. In a production setting, the output of the production process is observed, which leads to powerful methods to quantify deviations from optimizing behavior (called inefficiencies). In a consumption setting, we do not directly observe the utility obtained from consuming a bundle of goods. However, it is still possible the analyze the choices of individuals, on the basis of price and income information. In particular, the analysis of consumption choices is based on the behavioral hypothesis that a consumer chooses a bundle of goods that he prefers to all other bundles that he can afford. The analysis of consumption choices usually focuses on determining whether observed choices are rational or not. However, it is also possible to quantify deviations from optimizing behavior, in terms of money wasted by making suboptimal choices.

In my dissertation I have adopted a nonparametric approach to analyze production and consumption decisions. The advantage of a nonparametric approach is that there is no need to make (typically unverifiable) assumptions about the parametric specification of the production function or demand function, but rather 'lets the data speak for themselves'. A well-established nonparametric method to analyze producer behavior is Data Envelopment Analysis (DEA). The first part of my dissertation focused on production behavior and presents extensions and empirical applications of DEA-based methodologies. The second part of my dissertation discussed consumption behavior and employs revealed preference methodology, which is a popular nonparametric method to analyze consumption behavior. Let me briefly review the contributions set out in this dissertation, and subsequently discuss avenues for future research.

Contributions In Chapter 1 we highlighted the behavioral interpretation of DEA. The behavioral approach starts from a clear specification of the production-behavioral model - for example cost minimization or profit maximization - and imposes the least structure on the production possibilities. The appropriate behavioral model of course depends on the specific application. The behavioral approach is opposed to the conventional axiomatic approach for reconstructing production possibilities, which we apply later in Chapters 2 and 3. To recall, the axiomatic approach reconstructs the production possibility set on the basis of a number of production axioms, such as free disposability and convexity. The efficiency of a production unit is then measured as the distance of the observed input-output combination to the boundary of the production possibility set.

Interestingly, duality relationships link the behavioral to the axiomatic approach. We therefore argue that duality relationships can justify the use of particular production axioms. Still, we plead for carefully checking the validity of axioms and for investigating the sensitivity of the results with respect to these axioms if it is difficult to verify them empirically. In our empirical application on US universities, we illustrate that whether an axiom is imposed or not, can have a considerable impact on the efficiency results.

Chapters 2 and 3 extend the DEA-based methodology of Cherchye et al. (2013) for multi-output efficiency measurement. Multi-output efficiency measurement presents a more adequate view on production economics, compared to the standard production approach, by including additional information on the link between inputs and outputs. By using a multi-output approach, we obtain methods that have considerably more discriminatory power, which opens up new possibilities for further research. Moreover, having more detailed information on the sources of inefficiency, should lead to more targeted actions for efficiency improvement. Cherchye et al. (2013) explicitly include information about the decomposition of the inputs to the outputs, by distinguishing inputs that can be directly allocated to specific outputs (output-specific inputs) from inputs that simultaneously benefit the production of multiple outputs (joint inputs). This allows the authors to introduce output-specific production technologies, which remain linked through the use of joint inputs.

In Chapter 2, we use this input information to develop a method that investigates whether a reallocation of inputs across outputs can yield efficiency gains. In particular, we introduce the notion of coordination efficiency to quantify the gains from reallocation. Interestingly, our method also provides specific guidelines to achieve a more optimal input allocation.

We continue the empirical application of education and research at US universities. However, we now explicitly model the multi-output nature of US universities. We believe the multi-output methodology is particularly well-suited to analyze the joint production of education and research, as it can account for both joint inputs and inputs specifically allocated to education or research. We found that more than half of the universities in the sample could considerably improve their efficiency by adopting a more optimal input allocation (over education and research outputs).

In Chapter 3, we build on the multi-output methodology of Cherchye et al. (2013) to estimate output-specific economies of scale. The distinguishing feature of the methodology is that returns to scale may differ between the different dimensions of production. Furthermore, to estimate output-specific returns to scale, we take into account output-specific environmental heterogeneity. In particular, we allow that environmental variables may be relevant for the production of particular outputs, but not for others. Finally, the proposed methodology distinguishes between discretionary and non-discretionary output variables. We therefore measure the performance of DMUs only with respect to the output variables that the DMU management controls and actually wants to maximize. All the pieces are in place to perform a thorough examination of economies of scale using a complete multi-output assessment.

We employ the methodology to examine the optimal scale size of prisons in England and Wales, using publicly available data provided by the Ministry of Justice. Besides the task of incarcerating convicts, we consider in our study also qualitative outputs, including the provision of a humane prison environment and successful reintegration. Furthermore, we control for the inflow of prisoners in a particular prison and for the socioeconomic environment in which the prisoners released have to reintegrate. With respect to prisons in England and Wales, we do not find support for the idea that public managers are confronted with a prison size dilemma. Depending on the operating environment, we found that medium to large scale size is optimal for both incarceration and providing purposeful and outside-cell activities. For successful reintegration, the results are mixed, but we did not find indications for drastic productivity gains by moving towards a very small prison scale.

Finally, the methodology of Chapter 4 differs from the previous chapters. In Chapter 4, we have integrated household production in a revealed preference analysis of household consumption. In the household production function approach, first introduced by Becker (1965), households combine time and market goods to produce domestic commodities. The domestic commodities then generate utility for the households. We focus on the household production of child well-being, which is produced by means of parental time and consumption goods allocated to the children.

Furthermore, we include process benefits related to child care, by allowing that parents may enjoy spending time with the children. In our framework, child care time 'jointly' produces well-being for the household's children - modeled as household production -and leisure for the caring parent - modeled as a substitute for pure leisure. We therefore present a revealed preference test of a labor supply model with home production of child well-being and process benefits. A major advantage of the revealed preference methodology is that the method can deal with home production, without making functional assumptions about the household production technology. Furthermore, the corresponding output does not even have to be observed. This makes the methodology particularly useful to study the well-being of children.

The revealed preference test determines whether the observed choices (with respect to time use and consumption) are consistent with the labor supply model with home production and process benefits. If the choices are consistent, the behavior of the households is said to be rational. On the other hand, if households are not fully rational, we can quantify how much these households deviate from optimizing behavior, in terms of the money wasted by making sub-optimal choices. Towards this end, we introduce a measure of critical cost efficiency. A low critical cost efficiency indicates that the observed decisions are not in line with the theoretical model. We therefore use the critical cost efficiency as our goodness-of-fit measure, which quantifies if the model under study is a reasonable model to study the observed behavior. Note that we presented a test for the unitary framework and one for the collective framework. In the unitary framework, a household is assumed to act as a single decision maker. By contrast, collective models explicitly recognize that households consist of multiple decision makers, each having their own preferences. In the collective model, the choices of a household are assumed to be the result of a bargaining process.

We applied our test to a sample of Dutch households, retrieved from the Longitudinal Internet Studies for the Social sciences (LISS). Our results indicate that including process benefits considerably improves the goodness-of-fit of labor supply models with home production. As expected, we found that the goodness-of-fit of the collective model is higher than the unitary model. Still, we showed that the gains from including process benefits into the model were equally large (in terms of goodness-of-fit) for the unitary and the collective model. Moreover, we also discussed the discriminatory power of the test. Towards this end, we simulated random data sets and compared the goodness-of-fit of the observed choices with the (average) goodness-of-fit of the random choices. We found that the discriminatory power of the model with process benefits is much higher than the model without process benefits.

Further research In this dissertation I have discussed methods which are designed to test consistency of observed behavior with theoretically optimizing behavior. In my opinion a great challenge is to carefully specify the underlying behavioral model, to obtain a fair analysis of optimizing behavior. Taking all specificities of the production process into account not only leads to more realistic models, but also to methods with more discriminatory power. Also in the context of household consumption, it is crucial to specify behavioral models that fit the data as good as possible. At the same time, the corresponding tests of consistency with the model need to be capable of discriminating between observed and random behavior.

In this dissertation, we attempted to bridge the gap between consumption and production analysis. Chapter 4 was a first study to integrate household production in a revealed preference analysis of household consumption. I hope that the work in this dissertation may serve as a step up for many future research, which applies insights of the production setting to the consumption setting and vice versa. Let me outline some ideas for further research on the basis of the contributions in this dissertation.

First, note that it is perfectly possible to combine the methodological extensions in Chapter 2 and 3. In particular, we could include alternative returns to scale assumptions to investigate possible efficiency gains from input reallocation. The methodology in Chapter 2 currently assumes variable returns to scale. However, one could impose alternative returns to scale assumptions such as constant, increasing and decreasing returns to scale, which may differ from output to output. This would lead to a methodology which models the production process in a more realistic manner and has more power to identify efficiency gains from reallocation.

Second, the empirical application on economies of scale in prisons in England and Wales illustrates the potential of the multi-output methodology to provide guidance in policy debates. The methodology in Chapter 3 is tailored to all particularities of the prison production process and enables us to meaningfully answer the prison size dilemma. However, our research was based on publicly available data. The Ministry of Justice could refine the analysis by adding detailed information on the allocation of expenses to specific outputs. This would lead to an even more detailed insight into the multi-output production process of local male prisons.

Third, we focused in Chapter 3 on economies of scale in prisons. However, the methodology is more generally applicable and well suited to analyze economies of scale in multi-output public sector organizations and multi-output manufacturing plants. Just some examples are economies of scale in fire protection (Duncombe and Yinger (1993)), in education (e.g. Andrews et al. (2002) and Leithwood and Jantzi (2009)) and in health care (Given (1996) and Wholey et al. (1996)).

Furthermore, we showed in Chapter 4 that including process benefits into the model significantly improves the goodness-of-fit of the model for the observed data. In a next step, it would be interesting to link the magnitude of the process benefits to observed family characteristics. This information would also enable us to compare the level of process benefits we retrieved with the findings of studies that attempt to directly measure process benefits. Examples of such studies are Juster (1985), who asks in general whether people like an activity, and Krueger et al. (2009) who use a time use diary in which individuals can indicate the nature of the activity and the extent to which various emotions were present.

Next, Chapter 4 focused on a static model of consumption and time use. In particular, the level of child well-being only depends on current expenditures and current time use choices. An avenue for future research is to investigate how our framework extends to a more general intertemporal setting, in order to study human capital formation.

Finally, the role of fathers for the development the child has been widely discussed in the literature (see e.g. Yeung et al. (2001) and Lamb (2004)). It would be interesting to compare the effect of mother's time and father's time on the well-being of the child. In our framework, it is possible to impose additional structure on the elements underlying the households' decision problems. Structure on the domestic technology that 'produces' child well-being may shed light on the productivity of parental time inputs as well as on the substitutability between mothers' and fathers' time invested in children.

Bibliography

Afriat, S., 1972. Efficiency estimation of production functions. International Economic Review 13 (3), 568–598.

Afriat, S. N., 1967. The construction of utility functions from expenditure data. International Economic Review 8, 67–77.

Afriat, S. N., 1973. On a system of inequalities in demand analysis: An extension of the classical method. International Economic Review 14, 460–472.

Agasisti, T., 2013. The efficiency of italian secondary schools and the potential role of competition: a data envelopment analysis using oecd-pisa2006 data. Education Economics 21 (5), 520–544.

Ahn, T., Charnes, A., Cooper, W. W., 1988. Some statistical and dea evaluations of relative efficiencies of public and private institutions of higher learning. Socio-Economic Planning Sciences 22 (6), 259–269.

Andrews, M., Duncombe, W., Yinger, J., 2002. Revisiting economies of size in american education: are we any closer to a consensus? Economics of Education Review 21 (3), 245–262.

Ang, F., Kerstens, P. J., 2016. To mix or specialise? A coordination productivity indicator for English and Welsh farms. Journal of Agricultural Economics.

Apps, P. F., Rees, R., 1988. Taxation and the household. Journal of Public Economics 35, 355–369.

Atici, K. B., Podinovski, V. V., 2012. Mixed partial elasticities in constant returns-to-scale production technologies. European Journal of Operational Research 220 (1), 262–269.

Banker, R., Morey, R., 1986. Efficiency analysis for exogenously fixed inputs and outputs. Operations Research 34 (4), 513–521.

Banker, R. D., 1984. Estimating most productive scale size using data envelopment analysis. European Journal of Operational Research 17 (1), 35–44.

Banker, R. D., Charnes, A., Cooper, W. W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30 (9), pp. 1078–1092.

Banker, R. D., Cooper, W. W., Seiford, L. M., Thrall, R. M., Zhu, J., 2004. Returns to scale in different DEA models. European Journal of Operational Research 154 (2), 345–362.

Banker, R. D., Maindiratta, A., 1988. Nonparametric analysis of technical and allocative efficiencies in production. Econometrica 56 (6), 1315–1332.

Baumol, W., Panzar, J., Willig, R., 1982. Contestable markets and the theory of industry structure. Harcourt Brace Jovanovich.

Beasley, J., 1995. Determining Teaching and Research Efficiencies. The Journal of the Operational Research Society 46 (4), 441–452.

Becker, G. S., 1960. An economic analysis of fertility. In: Demographic and economic change in developed countries. Columbia University Press, pp. 209–240.

Becker, G. S., 1962. Irrational behavior and economic theory. Journal of Political Economy 70, 1–13.

Becker, G. S., 1965. A theory of the allocation of time. Economic Journal 75, 493–517.

Becker, G. S., 1968. Crime and punishment - economic approach. Journal of Political Economy 76 (2), 169–217.

Blundell, R., Chiappori, P., Meghir, C., 2005. Collective labour supply with children. Journal of Political Economy 105, 1277–1306.

Blundell, R., Robin, J.-M., 2000. Latent Separability: Grouping Goods Without Weak Separability. Econometrica 68 (1), 53–84.

Bogetoft, P., 1996. DEA on Relaxed Convexity. Management Science 42 (3), 457–465.

Bogetoft, P., 2004. Incentive efficient production frontiers: An agency perspective on dea. Management Science 40, 959–968.

Bonaccorsi, A., Daraio, C., Simar, L., 2006. Advanced indicators of productivity of universitiesan application of robust nonparametric methods to italian data. Scientometrics 66 (2), 389–410.

Bronars, S. G., 1987. The power of nonparametric tests of preference maximization. Econometrica 55, 693–698.

Browning, M., 1992. Children and household economic behavior. Journal of Economic Literature 30 (3), 1434–1475.

Bruyneel, S., Cherchye, L., De Rock, B., 2012. Collective consumption models with restricted bargaining weights: an empirical assessment based on experimental data. Review of Economics of the Household 10, 395–421.

Buonanno, P., Raphael, S., Oct. 2013. Incarceration and incapacitation: Evidence from the 2006 italian collective pardon. American Economic Review 103 (6), 2437–2465.

Camanho, A. S., Dyson, R. G., 2005. Cost efficiency measurement with price uncertainty: A DEA application to bank branch assessments. European Journal of Operational Research 161, 432–446.

Camanho, A. S., Dyson, R. G., 2008. A generalisation of the Farrell cost efficiency measure applicable to non-fully competitive settings. Omega: The International Journal of Management Science 36, 147–162.

Campbell, M. C., Vogel, M., Williams, J., May 2015. Historical contingencies and the evolving importance of race, violent crime, and region in explaining mass incarceration in the united states. Criminology 53 (2), 180–203.

Carter, P., 2007. Securing the future: Proposals for the efficient and sustainable use of custody in England and Wales. Her Majesty's Stationary Office, London. UK.

Cazals, C., Florens, J. P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. Journal Of Econometrics 106 (1), 1–25.

Chambers, R., Chung, Y., Färe, R., 1998. Profit, directional distance functions, and nerlovian efficiency. Journal of Optimization Theory and Applications 98, 351–364.

Charnes, A., Cooper, W. W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (429-444).

Cherchye, L., De Rock, B., Dierynck, B., Roodhooft, F., Sabbe, J., 2013. Opening the "black box" of efficiency measurement: Input allocation in multioutput settings. Operations Research 61 (5), 1148–1165.

Cherchye, L., De Rock, B., Estache, A., Verschelde, M., 2016. Efficiency measures in regulated industries: History, outstanding challenges and emerging solutions. In: E. Grifell-Tatje, C. A. K. L., Sickles, R. C. (Eds.), forthcoming in The Oxford Handbook of Productivity Analysis as Chapter 17. Oxford University Press, Oxford.

Cherchye, L., De Rock, B., Hennebel, V., 2014a. Coordination efficiency in multi-output settings: a dea approach. Annals of Operations Research, 1–29.

Cherchye, L., De Rock, B., Hennebel, V., 2014b. The economic meaning of data envelopment analysis: A 'behavioral'perspective. Socio-economic planning sciences 48 (1), 29–37.

Cherchye, L., De Rock, B., Vermeulen, F., 2007. The collective model of household consumption: a nonparametric characterization. Econometrica 75, 553–574.

Cherchye, L., De Rock, B., Vermeulen, F., 2008. Analyzing Cost-Efficient Production Behavior Under Economies of Scope : A Nonparametric Methodology. Operations Research 56 (1), 204–221.

Cherchye, L., De Rock, B., Vermeulen, F., 2011a. The revealed preference approach to collective consumption behavior: testing and sharing rule recovery. Review of Economic Studies 78, 176–198.

Cherchye, L., De Rock, B., Vermeulen, F., 2012. Married with children: A collective labor supply model with detailed time use and intrahousehold expenditure information. American Economic Review 102, 3377–3405.

Cherchye, L., De Rock, B., Vermeulen, F., 2015a. A simple identification strategy for gary becker's time allocation model. Economics Letters 137, 187–190.

Cherchye, L., De Rock, B., Walheer, B., 2015b. Multi-output efficiency with good and bad outputs. European Journal of Operational Research 240, 872–888.

Cherchye, L., De Witte, K., Ooghe, E., Nicaise, I., 2010a. Efficiency and equity in private and public education: A nonparametric comparison. European Journal of Operational Research 202 (2), 563–573.

Cherchye, L., Demuynck, T., De Rock, B., 2011b. Revealed preference analysis of noncooperative household consumption. Economic Journal 121, 1073–1096.

Cherchye, L., Demuynck, T., De Rock, B., De Witte, K., 2014c. Non-parametric analysis of multi-output production with joint inputs. Economic Journal 124 (577), 735–775.

Cherchye, L., Demuynck, T., De Rock, B., Hjertstrand, P., 2015c. Revealed preference tests for weak separability: An integer programming approach. Journal of Econometrics 186, 129–141.

Cherchye, L., Kuosmanen, T., Leleu, H., 2010b. Technical and economic efficiency measures under short run profit maximizing behavior. Louvain Economic Review 76, 163–173.

Cherchye, L., Kuosmanen, T., Post, T., 2000. What is the economic meaning of fdh? a reply to thrall. Journal of Productivity Analysis 13, 259–263.

Cherchye, L., Post, T., 2003. Methodological Advances in DEA: A survey and an application for the Dutch electricity sector. Statistica Neerlandica 57 (4), 410–438.

Cherchye, L., Van Puyenbroeck, T., 2007. Profit efficiency analysis under limited information. with an application to german farm types. OMEGA International Journal of Management Science 35, 204–221.

Chiappori, P., 1988. Rational household labor supply. Econometrica 56, 63–89.

Chiappori, P., 1992. Collective labor supply and welfare. Journal of Political Economy 100, 437–467.

Chiappori, P. A., 1997. Introducing household production in collective models of labor supply. Journal of Political Economy 105, 191–209.

Chiappori, P.-A., Lewbel, A., 2015. Gary becker's a theory of the allocation of time. The Economic Journal 125 (583), 410–442.

Cloïn, M., 2013. Met het oog op de tijd. Sociaal en Cultureel Planbureau.

Coelli, T., Estache, A., Perelman, S., Trujillo, L., 2003. A Primer on Efficiency Measurement for Utilities and Transport Regulators. Word Bank Institute.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., Battese, G. E., 2005. An introduction to efficiency and productivity analysis. Springer Science & Business Media.

Cook, W. D., Seiford, L. M., Jan. 2009. Data envelopment analysis (DEA) – Thirty years on. European Journal of Operational Research 192 (1), 1–17.

Cook, W. D., Zhu, J., 2011. Multiple Variable Proportionality in Data Envelopment Analysis. Operations Research 59 (4), 1024–1032.

Cook, W. D., Zhu, J. (Eds.), 2014. Data Envelopment Analysis: A Handbook of Modeling Internal Structure and Network. Springer.

Cooper, R., Kaplan, R. S., 1988. Measure costs right: Make the right decision. Harvard Bussiness Review 66 (5), 96–103.

Cooper, W. W., Seiford, L. M., Tone, K., 2007. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Springer.

Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: A probabilistic approach. Journal of Productivity Analysis 24 (1), 93–121.

Daraio, C., Simar, L., 2007a. Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. Studies in Productivity and Efficiency. Springer.

Daraio, C., Simar, L., 2007b. Conditional nonparametric frontier models for convex and nonconvex technologies : a unifying approach. Journal of Productivity Analysis 28, 13–32.

Dauphin, A., El Lahga, A. R., Fortin, B., Lacroix, G., 2011. Are Children Decision-Makers within the Household? Economic Journal 121 (553), 871–903.

De Witte, K., López-Torres, L., 2015. Efficiency in education: a review of literature and a way forward. Journal of the Operational Research Society.

De Witte, K., Rogge, N., 2010. To publish or not to publish? on the aggregation and drivers of research performance. Scientometrics 85 (3), 657–680.

De Witte, K., Rogge, N., Cherchye, L., Van Puyenbroeck, T., 2013. Economies of scope in research and teaching: a non-parametric investigation. Omega 41 (2), 305–314.

Debreu, G., 1951. The coefficient of resource utilization. Econometrica 19, 273–292.

Del Boca, D., 2002. The effect of child care and part time opportunities on participation and fertility decisions in italy. Journal of Population Economics 15 (3), 549–573.

Demsetz, H., 1997. The Economics of the Business Firm. Seven Critical Commentaries. Cambridge University Press.

Deprins, L., Simar, L., Tulkens, H., 1984. Measuring labor efficiency in post offices. In: Marchand, M., Pestieau, P., Tulkens, H. (Eds.), The Performance of Public Enterprises: Concepts and Measurement. North Holland, Amsterdam, pp. 243–267.

Di Tella, R., Schargrodsky, E., Feb. 2013. Criminal recidivism after prison and electronic monitoring. Journal of Political Economy 121 (1), 28–73.

Diewert, W. E., 1973. Afriat and revealed preference theory. The Review of Economic Studies 40, 419–425.

Diewert, W. E., Parkan, C., 1983. Linear programming tests of regularity conditions for production frontiers. In: W. Eichhorn, R. Henn, K. N., Shepard, R. (Eds.), Quantitative studies in production and prices. Physica-Verlag, Würzburg, pp. 131–158.

Dow, G. K., Juster, F. T., 1985. Goods, time and well-being: The joint dependence problem. Time, goods and well-being, 397–413.

Duncombe, W., Miner, J., Ruggiero, J., 1997. Empirical evaluation of bureaucratic models of inefficiency. Public Choice 93 (1-2), 1–18.

Duncombe, W., Yinger, J., 1993. An analysis of returns to scale in public production, with an application to fire protection. Journal of Public Economics 52 (1), 49–72.

Emrouznejad, A., Parker, B., Tavares, G., 2008. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in dea. Socio-Economic Planning Sciences 42, 151–157.

Färe, R., Grabowski, R., Grosskopf, S., Kraft, S., 1997. Efficiency of a fixed but allocatable input : A non-parametric approach. Economics Letters 56, 187–193.

Färe, R., Grosskopf, S., 1994. Cost and Revenue Constrained Production. Springer.

Färe, R., Grosskopf, S., 1995. Non-parametric tests of regularity, farrell efficiency, and goodness of fit. Journal of Econometrics 69, 415–425.

Färe, R., Grosskopf, S., 2000. Network DEA. Socio-Economic Planning Sciences 34, 35–49.

Färe, R., Grosskopf, S., Lovell, C., 1994. Production Frontiers. Cambridge University Press.

Färe, R., Grosskopf, S., Nelson, J., 1990. On price efficiency. International Economic Review 31, 709–720.

Färe, R., Grosskopf, S., Whittaker, G., 2007. Network dea. In: Modeling data irregularities and structural complexities in data envelopment analysis. Springer, pp. 209–240.

Farrell, M. J., 1957. The Measurement of Productive Efficiency. Journal of the Royal Statistical Society. Series A (General) 120 (3), pp. 253–290.

Ferrera, J. M. C., Cebada, E. C., Chaparro, F. P., González, D. S., 2011. Exploring educational efficiency divergences across spanish regions in pisa 2006. Revista de economía aplicada 19 (57), 117–146.

Fried, H., Lovell, C., Schmidt, S., 2008. The measurement of productive efficiency and productivity growth. Oxford University Press, Oxford.

Fukuyama, H., Weber, W. L., 2008. Profit inefficiency of Japanese securities firms. Journal of Applied Economics 11, 281–303.

Gattoufi, S., Oral, M., Reisman, A., 2004. Data envelopment analysis literature: A bibliography update (1951-2001). Journal of Socio-Economic Planning Sciences 38, 159–229.

Gershuny, J., Halpin, B., 1996. Time use, quality of life and process benefits. In pursuit of the quality of life, 188–210.

Given, R. S., 1996. Economies of scale and scope as an explanation of merger and output diversification activities in the health maintenance organization industry. Journal of health economics 15 (6), 685–713.

Gørtz, M., 2011. Home production - enjoying the process or the product? International Journal of Time Use Research 8, 85–109.

Graham, J. W., Green, C. A., 1984. Estimating the parameters of a household production function with joint products. Review of Economics and Statistics 66, 277–282.

Gronau, R., 1977. Leisure, home production, and work: the theory of the allocation of time revisited. Journal of Political Economy 85, 1099–1124.

Grosskopf, S., 1996. Statistical inference and non-parametric efficiency: A selective survey. Journal of Productivity Analysis 7, 161–176.

Grosskopf, S., 2003. Some remarks on productivity and its decompositions. Journal of Productivity Analysis 20 (3), 459–474.

Guryan, J., Hurst, E., Kearny, M. S., 2008. Parental education and parental time with children. Journal of Economic Perspectives 22, 23–46.

Hall, R. E., 1973. The specification of technology with several kinds of output. The Journal of Political Economy, 878–892.

Hallberg, D., Klevmarken, A., 2003. Time for children: A study of parent's time allocation. Journal of Population Economics 16, 205–226.

Hanoch, G., Rothschild, M., 1972. Testing the assumptions of production theory: A nonparametric approach. Journal of Political Economy 80 (2), pp. 256–275.

Hansen, B., Apr. 2015. Punishment and deterrence: Evidence from drunk driving. American Economic Review 105 (4), 1581–1617.

HM Inspectorate of Prisons, 2009. The prison characteristics that predict prisons being assessed as performing 'well': A thematic review. Tech. rep., HM Chief Inspector of Prisons, London: HM Inspectorate of Prisons.

Houthakker, H. S., 1950. Revealed preference and the utility function. Economica 17 (66), pp. 159–174.

Juster, F. T., 1985. Preferences for work and leisure. Time, goods, and well-being, 335–51.

Juster, F. T., Courant, P. N., Dow, G. K., 1985. A conceptual framework for the analysis of time allocation data. Time, goods, and well-being, 113–31.

Kalb, G., 2009. Children, labour supply and child care: challenges for empirical analysis. Australian Economic Review 42 (3), 276–299.

Katz, L., Levitt, S., Shustorovich, E., 2003. Prison conditions, capital punishment, and deterrence. American Law and Economics Review 5 (2), 318–343.

Kerkhofs, M., Kooreman, P., 2003. Identification and estimation of a class of household production models. Journal of Applied Econometrics 18, 337–369.

Kerstens, K., Vanden Eeckaut, P., 1998. Distinguising Technical and Scale Efficiency on Non-Convex and Convex Technologies: Theoretical Analysis and Empirical Illustrations. CORE discussion paper 98.55.

Kerstens, K., Vanden Eeckaut, P., 1999. Estimating returns to scale using non-parametric deterministic technologies: A new method based on goodness-of-fit. European Journal of Operational Research 113 (1), 206–214.

Kohli, U., 1983. Non-joint technologies. The Review of Economic Studies 50 (1), 209–219.

Kohli, U., 1985. Technology and public goods. Journal of Public Economics 26 (3), 379–400.

Krueger, A. B., Kahneman, D., Schkade, D., Schwarz, N., Stone, A. A., 2009. National time accounting: The currency of life. In: Measuring the subjective well-being of nations: National accounts of time use and well-being. University of Chicago Press, pp. 9–86.

Kuosmanen, T., Post, T., 2002. Nonparametric efficiency analysis under uncertainty: A first-order stochastic dominance approach. Journal of Productivity Analysis 17, 183–200.

Lamb, M. E., 2004. The role of the father in child development. John Wiley & Sons.

Lau, L. J., 1972. Profit functions of technologies with multiple inputs and outputs. The Review of Economics and Statistics, 281–289.

Leithwood, K., Jantzi, D., 2009. A review of empirical evidence about school size effects: A policy perspective. Review of educational research 79 (1), 464–490.

Levitt, S. D., May 1996. The effect of prison population size on crime rates: Evidence from prison overcrowding litigation. Quarterly Journal of Economics 111 (2), 319–351.

Liebling, A., 2004. Prisons and Their Moral Performance: A Study of Values, Quality, and Prison Life. Oxford: Oxford University Press.

Lise, J., Yamada, K., 2014. Household sharing and commitment: Evidence from panel data on individual expenditures and time use. Tech. rep., Institute for Fiscal Studies.

Lockyer, K., 2013. A radical plan to reform the prison estate. Policy exchange, London. UK.

Lozano, S., 2015. A joint-inputs network dea approach to production and pollution-generating technologies. Expert Systems with Applications 42 (21), 7960–7968.

Lozano, S., Villa, G., Jul. 2004. Centralized Resource Allocation Using Data Envelopment Analysis. Journal of Productivity Analysis 22 (1/2), 143–161.

Mayston, D. J., 2015. Data envelopment analysis, endogeneity and the quality frontier for public services. Annals of Operations Research, 1–19.

McCorkle, R. C., Miethe, T. D., Drass, K. A., Jul. 1995. The roots of prison violence - a test of the deprivation, management, and not-so-total institution models. Crime & Delinquency 41 (3), 317–331.

McFadden, D., 1978. Cost, revenue, and profit functions. In: Fuss, M., McFadden, D. (Eds.), Production Economics: A Dual Approach to Theory and Applications. North-Holland, Amsterdam, pp. 3–110.

Mears, D. P., Stewart, E. A., Siennick, S. E., Simons, R. L., Aug. 2013. The code of the street and inmate violence: Investigating the salience of imported belief systems. Criminology 51 (3), 695–728.

Ministry of Justice, 2011. Breaking the cycle: Government response. London: The Stationery Office.

Ministry of Justice, 2013. Transforming Rehabilitation: A strategy for reform. London: The Stationery Office.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., Giovannini, E., 2008. Handbook on constructing composite indicators: Methodology and User guide. OECD publishing.

National Audit Office, 2013. Managing the prison estate. Tech. rep., London: The Stationary Office.

Nerlove, M., 1965. Estimation and identification of Cobb-Douglas production functions. Rand McNally.

Niskanen, W. A., 1971. Bureaucracy and representative government. Transaction Publishers.

Olesen, O. B., Petersen, N. C., Podinovski, V. V., 2015. Efficiency analysis with ratio measures. European Journal of Operational Research 245 (2), 446–462.

Petersen, N. C., 1990. Data Envelopment Analysis on a Relaxed Set of Assumptions. Management Science 36 (3), 305–314.

Podinovski, V. V., 2004a. Efficiency and Global Scale Characteristics on the " No Free Lunch " Assumption Only. Journal of Productivity Analysis 22, 227–257.

Podinovski, V. V., 2004b. Local and Global Returns to Scale in Performance Measurement. Journal of the Operational Research Society 55 (2), 170–178.

Podinovski, V. V., 2004c. On the linearisation of reference technologies for testing returns to scale in fdh models. European Journal of Operational Research 152 (3), 800–802.

Podinovski, V. V., Forsund, F. R., 2010. Differential characteristics of efficient frontiers in data envelopment analysis. Operations Research 58 (6), 1743–1754.

Podinovski, V. V., Kuosmanen, T., 2011. Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions. European Journal of Operational Research 211 (3), 577–585.

Pollak, R. A., Wachter, M. L., 1975. The relevance of the household production function and its implications for the allocation of time. Journal of Political Economy 83, 255–278.

Reny, P. J., 2015. A characterization of rationalizable consumer behavior. Econometrica 83 (1), 175–192.

Robinson, J., Godbey, G., 1997. Time for life: The surprising ways Americans use their time. Penn State Press.

Rogge, N., Simper, R., Verschelde, M., Hall, M., 2015. An analysis of managerialism and performance in English and Welsh male prisons. European Journal of Operational Research 241 (1), 224–235.

Ruggiero, J., 1996. On the measurement of technical efficiency in the public sector. European Journal of Operational Research 90 (3), 553–565.

Ruggiero, J., 2000. Nonparametric estimation of returns to scale in the public sector with an application to the provision of educational services. Journal of the Operational Research Society 51 (8), 906–912.

Ruiz, C. F., Bonilla, R., Chavarro, D., Orozco, L. A., Zarama, R., Polanco, X., 2010. Efficiency measurement of research groups using data envelopment analysis and bayesian networks. Scientometrics 83 (3), 711–721.

Sahoo, B. K., Mehdiloozad, M., Tone, K., 2014. Cost, Revenue and Profit efficiency Measurement in DEA: A Directional Distance Function Approach. European Journal of Operational Research 237, 921–931.

Sahoo, B. K., Tone, K., Jan. 2013. Non-parametric measurement of economies of scale and scope in non-competitive environment with price uncertainty. Omega 41 (1), 97–111.

Samuelson, P. A., 1938. A note on the pure theory of consumer's behavior. Economica 5, 61–71.

Samuelson, P. A., 1948. Consumption theory in terms of revealed preference. Economica 15 (60), pp. 243–253.

Samuelson, P. A., 1966. The fundamental singularity theorem for non-joint production. International Economic Review 7 (1), 34–41.

Schwarz, N., Kahneman, D., Xu, J., Belli, R., Stafford, F., Alwin, D., 2009. Global and episodic reports of hedonic experience. Using calendar and diary methods in life events research, 157–174.

Shepard, R., 1970. Theory of Cost and Production Functions. Princeton University Press, Princeton.

Simar, L., Wilson, P., 2007. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 136 (1), 31–64.

Sykes, G., 1958. The society of captives: A study of a maximum security prison. Princeton, NJ: Princeton University Press.

Syrjänen, M. J., 2004. Non-discretionary and discretionary factors and scale in data envelopment analysis. European Journal of Operational Research 158 (1), 20–33.

Tone, K., Sahoo, B. K., May 2003. Scale, indivisibilities and production function in data envelopment analysis. International Journal of Production Economics 84 (2), 165–192.

Tulkens, H., 1993. On fdh efficiency analysis: Some methodological issues and applications to retail banking, courts and urban transit. Journal of Productivity Analysis 4, 183–190.

Tulkens, H., Vanden Eeckaut, P., 1999. Mésurer l'efficacité: avec ou sans frontiéres? In: Badillo, P.-Y., Paradi, J. (Eds.), La Méthode DEA : Analyse des Performances. Hermes Science, Paris, pp. 75–100.

van den Heuvel, P., 1986. Nonjoint production and the cost function: some refinements. Journal of Economics 46 (3), 283–297.

Varian, H. R., 1982. The nonparametric approach to demand analysis. Econometrica 50, 945–973.

Varian, H. R., 1983. Non-parametric Tests of Consumer Behaviour. The review of economic studies 50 (1), 99–110.

Varian, H. R., 1984. The Nonparametric Approach to Production Analysis. Econometrica 52 (3), 579–597.

Varian, H. R., 1990. Goodness-of-fit in optimizing models. Journal of Econometrics 46, 125–140.

Vinkler, P., 1998. General performance indexes calculated for research institutes of the hungarian academy of sciences based on scientometric indicators. Scientometrics 41 (1-2), 185–200.

Vinkler, P., 2006. Composite scientometric indicators for evaluating publications of research institutes. Scientometrics 68 (3), 629–642.

Vollaard, B., Mar. 2013. Preventing crime through selective incapacitation. Economic Journal 123 (567), 262–284.

Wholey, D., Feldman, R., Christianson, J. B., Engberg, J., 1996. Scale and scope economies among health maintenance organizations. Journal of health economics 15 (6), 657–684.

Yeung, W. J., Sandberg, J. F., Davis-Kean, P. E., Hofferth, S. L., 2001. Children's time with fathers in intact families. Journal of Marriage and Family 63 (1), 136–154.

List of doctoral dissertations

Doctoral dissertations from the Faculty of Business and Economics, see:

http://www.kuleuven.ac.be/doctoraatsverdediging/archief.htm